Groupthink:

Collective Delusions in Organizations and Markets

Roland Bénabou

Princeton University

First draft: May 2007

This version: August 2008

Abstract

I develop a model of (individually rational) *collective reality denial* in groups and organizations, or among participants in a market. More generally, I ask when individual tendencies toward wishful thinking and overoptimism reinforce or dampen each other.

To make clear that such groupthink is entirely distinct from standard linkage mechanisms, there are no complementarities or substitutabilities in payoffs in the basic model, nor any private signals that could give rise to social learning or herding. What emerges is thus a new and surprisingly simple mechanism generating interdependencies in information processing, beliefs and actions. Intuitively, whenever an agent benefits (on average) from other' delusions, this tends to make him more of a realist; and whenever their disconnect from reality makes him worse off this pushes him toward denial, which is then contagious. This *Mutually Assured Delusion* (MAD) principle can, in particular, give rise to multiple equilibria with different "social cognitions" of the same reality.

The same general principle implies that, in organizations where some agents have a greater impact on others' welfare than the reverse (e.g., managers and workers respectively), strategies of realism or denial will "trickle down" the hierarchy, so that subordinates will in effect *take their beliefs from the leader(s).* In addition to collective illusions of control, the model also accounts for the mirror case of collective fatalism and resignation, such as public apathy and "looking away" from humanitarian disasters.

In market interactions, equilibrium prices typically introduce a substitutability between agents' decisions that works against collective belief. Nonetheless, I show how, in markets with time-to build features, or more generally where participants find themselves with outstanding positions potentially subject to (endogenous) capital losses, contagious wishful thinking can again take hold, leading to overinvestment and an ultimate crash.

Finally, the model's welfare analysis makes clear what factors distinguish valuable group morale from harmful groupthink and generates new results concerning organizations' ex ante and ex post attitudes toward dissenting speech.

"It appears that there are enormous differences of opinion as to the probability of a failure with loss of vehicle and of human life. The estimates range from roughly 1 in 100 to 1 in 100,000. The higher figures come from the working engineers, and the very low figures from management. What are the causes and consequences of this lack of agreement? Since 1 part in 100,000 would imply that one could put a Shuttle up each day for 300 years expecting to lose only one, we could properly ask 'What is the cause of management's fantastic faith in the machinery?' "   (Richard Feynman, in *Rogers Commission Report,* 1986)

"We have a wealth of information we didn't have before," Joe Anderson, then a senior Countrywide executive, said in a 2005 interview. "We understand the data and can price that risk."   (BusinessWeek, "Not So Smart," August. 2007)

Introduction

This paper examines how collective beliefs and delusions arise and persist in organizations such as teams, firms, bureaucracies or markets. In the aftermath of corporate and public-policy disasters, it indeed often emerges that participants fell prey to a collective form of overconfidence and willful blindness: clear warning signals were systematically ignored or met with denial, evidence avoided, cast aside or selectively reinterpreted, dissenters discouraged and shunned. Market bubbles and manias exhibit the same pattern of investors acting "color-blind in a sea of red flags", followed by a crash (see Shiller (2005) for numerous examples).[1]

Janis (1972), analyzing policy decisions such as the Bay of Pigs invasion, the Cuban missile crisis and the escalation of the Vietnam war, identified in those that ended disastrously a cluster of such symptoms for which he coined the term "groupthink". Although some later work was critical of his characterization of those episodes, the concept has flourished and spurred a large literature in social and organizational psychology. Defined in Merriam-Webster's dictionary as *"a pattern of thought characterized by self-deception, forced manufacture of consent, and conformity to group values and ethics",* groupthink was strikingly documented in the official inquiries conducted on the Challenger and Columbia space shuttle disasters. It has also been invoked as a contributing factor in the failures of companies such as Enron and Worldcom, in some decisions relating to the second Iraq war, and most recently in the subprime-loan and housing-market crisis.[2] At the same time, one should keep in mind that the mirror opposite of harmful "groupthink" is precious "group morale" and seek to understand how they differ, even though both involve the maintenance of collective optimism in spite of negative signals.

---

[1] I borrow here the evocative title of Norris' (2008) account of mortgage securitization at Merrill Lynch.

[2] On the space shuttle accidents, see Rogers Commission (1986) and Columbia Accident Investigation Board (2003). On Enron, see Samuelson (2001), Barrienuovo (2006), Brunswick and Hayes (2002), Cohan (2002), Eichenwald (2003, 2005) and Pearlstein (2006). On Iraq, see Isikoff and Corn (2002), Hersh (2004), Suskind (2004) and Kaplan (2008). On self-deception and self-rationalizations as key enablers of corporate misconduct, see Huseman and Driver (1979), Sims (1992), Tenbrusel and Messick (2004) and Anand, Ashforth and Joshi (2005). On housing and other bubbles see Shiller (2005), and on the subprime mortgage crisis see Morgenson and Fabrikant (2007) and Andrews (2007).

To analyze these issues, I develop a model of (individually rational) *collective reality denial* in groups and organizations engaged in a joint project, or among participants in a market. The model, which builds on and extends the selective-awareness (attention, memory) framework of Bénabou and Tirole (2002, 2006a), allows me to ask when individual tendencies toward wishful thinking and overoptimism reinforce or dampen each other. To make clear that groupthink is entirely distinct from standard linkage mechanisms, there are *no complementarities or substitutabilities* in payoffs in the basic model, nor any private signals that could give rise to social learning or herding. What emerges is thus a new and surprisingly simple mechanism generating interdependencies in information processing, beliefs and actions. Intuitively, whenever an agent benefits (on average) from other' delusions, this tends to make him more of a realist; and whenever their disconnect from reality makes him worse off this pushes him toward denial, which is then contagious. This *Mutually Assured Delusion* (MAD) principle can, in particular, give rise to multiple equilibria with different "social cognitions" of the same reality.

The same general principle implies that, in organizations where some agents have a greater impact on others' welfare than the reverse (e.g., managers and workers respectively), strategies of realism or denial will "trickle down" the hierarchy, so that subordinates will in effect *take their beliefs from the leader(s).* In addition to collective illusions of control, it can also account for the mirror case of collective fatalism and resignation, such as public apathy and "looking away" from humanitarian disasters.

The model's welfare analysis makes clear what factors distinguish valuable group morale from harmful groupthink, and leads to interesting results concerning attitudes toward dissenting speech. In particular, it shows why organizations and societies can find it desirable to set up *ex-ante* commitment mechanisms protecting and encouraging dissent (constitutional guarantees of free speech, whistle-blower protections, devils' advocates, etc.), even when *ex-post* everyone would unanimously want to ignore or "kill" the messengers of bad news.

Turning finally to market interactions, prices typically introduce a substitutability between agents' decisions that works against collective belief. Nonetheless, I show how, in markets with time-to build features, or more generally where participants find themselves with substantial inventories or outstanding positions potentially subject to (endogenous) capital losses, *contagious wishful thinking* can again take hold, leading to overinvestment and an ultimate crash.

Related Literature. This work relates to four strands of literature. The first centers on cognitive dissonance, self deception and belief distortion more generally.[3] The second, closely related, is that on anticipatory utility.[4] The concern with group morale and groupthink in

---

[3]See, e.g., Akerlof and Dickens (1982), Schelling (1986), Kuran (1993), Rabin (1994), Carrillo and Mariotti (2000) and Bénabou and Tirole (2002, 2004, 2006a).

[4]See Loewenstein (1987), Caplin and Leahy (2001, 2003), Landier (2000), Caplin and Eliaz (2005), Brunnermeier and Parker (2005), Bernheim and Thomadsen (2005), Köszegi (2006, 2007), or Bénabou and Tirole (2007).

organizations ties into recent work on overoptimism and heterogeneous beliefs in firms.[5] Finally, the application to market manias and crashes links the paper to the literatures on bubbles, fads and herding, although the mechanism identified here is completely different.[6]. Beyond economics, the paper relates to the large literature in psychology on motivated beliefs and wishful thinking, to the organizational literature on groupthink-like phenomena (particularly corporate disasters) that followed Janis' (1972) book, and to the work in social and political psychology on "social cognition".[7]

1   Groupthink in teams and organizations

1.1   The basic model

Technology. A group of risk-neutral agents, $i \in \{1,...n\}$, are engaged in a joint project (team, firm, military unit) or other activities generating a public good or spillovers. At $t = 1$, each chooses an effort level $e^i = 0$ or 1, with cost $ce^i, c > 0$. At $t = 2$, he will reap expected utility

$$(1) \qquad\qquad U_2^i \equiv \theta \left[\alpha e^i + (1 - \alpha) e^{-i} \right],$$

where e^{-i} denotes the average effort of others,

$$(2) \qquad\qquad e^{-i} \equiv \frac{1}{n-1} \sum_{j \neq i} e^j,$$

and $1 - \alpha \in [0, 1 - 1/n]$ the degree of interdependence between agents, arising from the collective nature of the activity, or from cross-investments. Depending on α, the choice of e^i thus ranges from a pure private good (or bad) to a pure public one.[8] This linear payoff structure is maximally simple: all agents play symmetric roles, there is a fixed value to inaction $e = 0$, normalized to 0, and *no complementarity or interdependence of any kind* between agents' effort decisions.[9] These assumptions serve only to highlight the key mechanism, and will be relaxed later on.

---

[5]On the theoretical side, see, e.g. Rotemberg and Saloner (1993), Bénabou and Tirole (2003), Fang and Moscarini (2005), Gervais and Goldtsein (2005), Van den Steen (2005) and Landier, Sraer and Thesmar (2006). On the empirical side and focussing on CEO overconfidence, see, e.g. Malmendier and Tate (2005, 2008) or Camerer and Malmendier (2007).

[6]See, e.g., Bikhchandani, Hirshleifer and Welch (1998), Banerjee (1992), Caplin and Leahy (1994) or Scheinkman and Xiong (2003).

[7]For introductions to the organizational literature on groupthink, see, e.g. Huseman and Driver (1979) and Haslam (2004). On social cognition, see, e.g. Lerner's (1982) "belief in a just world", Jost and Major's (2001), "system justification", Leung et al. (2002)(), "social axioms", or Kahan et al.'s (2006) "cultural cognition".

[8] Another source of interdependence arises from *altruistic concerns* among agents: family or kinship ties, valued team identity, etc. Thus, (1) is equivalent to $U_2^i \equiv \beta \theta e^i + (1 - \beta) U_2^{-i}$ with $1 - \alpha \equiv (1 - \beta)(n - 1)/(n - \beta)$. Note also that while the notation in (1) suggests constant returns (e.g., "publicly provided private goods"), any crowding or scale economies can be reflected by dividing θ by some appropriate function of n.

[9]I intentionally abstract from complementarities and substitutabilities to demonstrate that they are neither necessary nor sufficient for groupthink, which, at its core, involves only the interplay of *cognitive* decisions.

The overall productivity of the venture agents are engaged in is a priori uncertain: $\theta = \theta_H$ in state H (probability q) and $\theta = \theta_L$ in state L (probability $1 - q$), with $\Delta\theta \equiv \theta_H - \theta_L > 0$ and $\theta_H > 0$ without loss of generality. Depending on the context, θ can represent the potential value of a firm's product or business plan, the state of the market, the suitability of a political or military strategy, or the quality of a leader. Note that θ also corresponds to the expected social value for the group of a choice $e^j = 1$, *relative* to what the alternative course of action would yield; the private value to the individual is $\alpha\theta - c$. If $\theta_L \geq 0$, each agent would always prefer that others choose $e = 1$ (put effort into a team project rather than rest, refrain from polluting, etc. If $\theta_L < 0$, however, he would like them to pursue the "appropriate" course of action for the organization, choosing $e = 1$ in state H and $e = 0$ in state L.

Preferences. The flow payoffs received by an agent during period 1 include the cost of effort, $-ce^i$, plus the *anticipatory utility* experienced from thinking about his future prospects, $sE_1^i[U_2^i]$, where $s \geq 0$ parametrizes the importance of hope, anxiety, dread, and similar emotions.[10] This parameter (s stands for "savoring" or "susceptibility") typically increases with the length of period 1, during which uncertainty remains; it may also vary across individuals.

At the start of period 1, agent i chooses effort to maximize the discounted value of payoffs,

$$(3) \qquad U_1^i = -ce^i + sE_1^i[U_2^i] + \delta E_1^i[U_2^i].$$

Given (1), his effort is determined solely by his beliefs about θ : $e^i = 1$ if $(s + \delta)\alpha E_1^i[\theta] > c$, *independently* of what any one else may be doing. I shall assume that

$$(4) \qquad \theta_L < \frac{c}{(s+\delta)\alpha} < \frac{c}{\delta\alpha} < q\theta_H + (1-q)\theta_L.$$

Thus, absent credible information, an individual acting on his prior will choose $e^i = 1$, whereas one who knows for sure that the state is L will abstain.[11]

An agent's beliefs at $t = 1$ depend on the news received at $t = 0$ and on how he processed them –accepting reality or averting his eyes from it, as specified below. In doing so, he acts so as to maximize the discounted utility of all payoffs,

$$(5) \qquad U_0^i = -M + \delta E_0^i\left[-ce^i + sE_1^i[U_2^i]\right] + \delta^2 E_0^i[U_2^i],$$

where M represents the date-0 costs, if any, of his information-processing strategy. The tradeoff between holding accurate versus hopeful beliefs embodied in these preferences will manifest itself

---

[10]This includes the well-documented health effects of (say) chronic stress versus hopefulness. For models of anticipatory utility under uncertainty see, e.g., Caplin and Leahy (2001), Köszegi (2005), Brunnermeier and Parker (2005), Bénabou and Tirole (2007), and Brunnermeier, Gollier and Parker (2007). The linear specification, $sE_1^i[U_2^i]$, avoids exogenously building into the model either information aversion or information-loving.

[11]This assumption is not essential but will ensure that each agent has a unique best-response awareness strategy, given that of others; see footnote 17 for details.

in agents' behavior with respect to both date-0 information and date-1 choices.

Information and beliefs. To represent agents' cognitive decisions or tendencies, I use a variant of the recall or awareness "technology" introduced in Bénabou and Tirole (2002, 2006a). At $t = 0$, agents observe a common signal that defines the relevant state of the world: $\sigma = H, L$, with probabilities q and $1 - q$ respectively.[12] Each one then has some flexibility in how much attention to pay to it, how to interpret it, whether to "keep it in mind" or "not think about it", etc. Formally, he can :

(a) Accept the news realistically, thus truthfully encoding $\hat{\sigma}^i = \sigma$ into memory or awareness (his date-1 information set).

(b) Engage in denial, censoring or rationalization, thus encoding $\hat{\sigma}^i = H$ instead of $\sigma = L$, or $\hat{\sigma}^i = L$ instead of $\sigma = H$. In addition to impacting later decisions, this may entail an immediate cost $m \geq 0$.[13]

(c) Deal in partial truths, using a mixed strategy. Equivalently, the memory process itself can be stochastic, with any recall probability $\lambda \in [0, 1]$ achievable at cost $M = m(1 - \lambda)$.

This simple informational structure captures a broad range of situations. For instance, the prior distribution $(q, 1 - q)$ could itself be conditional on some other signal being good news, such as the appearance of a new technology or market opportunity (versus a status quo where θ is low for sure). This positive signal may also have warranted some initial investment in the activity, including the formation of the group itself. Alternatively, it could contemporaneous to the realization of σ; L is then a state of "mixed evidence", whereas in H all signals are "go".

Directed attention and inattention. Instead of "tuning out" unwelcome news (denial), selective awareness can also take the form or investing extra resources in retaining good ones (rehearsal, preserving evidence). This corresponds to the case where attention or recall is naturally imperfect ($\lambda < 1$) but can be raised at some cost (it is like setting $m < 0$ in (b) above). Both mechanisms lead to broadly similar results and can be combined: what matters is that there be a possibility (and a motive) for *differential awareness* of H and L, not how this is achieved. While costly recall may be a more familiar assumption, actual episodes of groupthink, market manias, etc., typically involve the more striking phenomena of willful inattention, ex-post rationalizations, refusals to face the evidence, silencing of doubters and similar forms of information disregard. For this reason, the model emphasizes "selective inattention" more than

---

[12] Since θ_H or θ_L is only the *expected* value of the project conditional on σ, a low signal does not preclude a high final realization, and vice versa. The perfect correlation of signals across individuals is chosen for simplicity (it just needs to positive) and to make clear that the mechanism at work here has nothing to do with herding or informational cascades, in which agents with private signals make inferences from each other's behavior.

[13] Self-deception may be a conscious decision or an unconscious tendency, and the resources expended in the process may be material (eliminating evidence, avoiding certain people or situations, searching for and rehearsing more desirable signals) or mental ones (stress from repression, cognitive dissonance, guilt). As discussed below, any arbitrarily small $m > 0$ suffices to rule out uninteresting "babbling" equilibria in which there is censoring in both states ($\lambda_L < 1, \lambda_H < 1$). Beyond this, all the paper's key results apply equally well with $m = 0$, though non-zero costs are more realistic (particularly for the welfare analysis).

"selective attention".[14]

A first result is that, no matter how small $m > 0$, an agent will never censor signals in both states: either $\lambda_H = 1$, or $\lambda_L = 1$. Given (1), moreover, intuition suggests that it is only in the "bad-news" state L that he may do so: agents with anticipatory utility would not want to substitute bad news for good ones.[15] Verifying these claims in the appendix (Lemma 4.1), I focus for the time being on cognitive decisions in state L, denoted simply

$$(6) \qquad \lambda \equiv \Pr\left[\hat\sigma = L | \sigma = L\right].$$

Later on I will consider payoffs structures more general than (1), under which either state may (endogenously) be censored.

While agents can selectively process information, their latitude to affect beliefs remains constrained by Bayesian rationality: at $t = 1$, agent i may no longer have direct access to the original signal, but if he (as others) has a systematic tendency toward selective attention or interpretation, he will take that into account, using Bayes' rule to form posteriors. Thus, when $\hat\sigma^i = L$ the agent knows that the state is L, but when $\hat\sigma^i = H$ his posterior belief is only

$$(7) \qquad \Pr\left[\sigma = H \mid \hat\sigma^i = H, \lambda^i\right] = \frac{q}{q + (1-q)(1-\lambda^i)} \equiv r(\lambda^i),$$

where λ^i is is his *equilibrium* rate of realism.[16]

To analyze the Perfect Bayesian equilibria of this game, I proceed in three steps. First, I fix everyone but agent i's awareness strategy at some arbitrary $\lambda^{-i} \in [0, 1]$ and look for his "best response" λ^i.[17] Second, I identify the general principle that governs whether individual cognitions are *substitutes* (the more others delude themselves, the better informed I want to be) or *complements* (the more others delude themselves, the less I also want to face the truth). Finally, I derive conditions under which groupthink arises in its most striking form, where both collective realism and collective denial constitute self-sustaining *social cognitions*.

---

[14] This also eliminates some uninteresting technicalities that arise when all signals are costly to retain. In the macroeconomic literature on "rational inattention" (e.g., Sims (2003), Reis (2006)), agents face ex-ante costs of (or limits on) information acquisition and processing. Here, by contrast, they make (equally rational) ex-post choices about how to deal with information they have been exposed to. Combining costly ex-ante attention decisions with anticipatory utility or some other source of motivated beliefs (thus linking the two literatures) would leave the main insights unchanged.

[15] An agent who likes pleasant surprises and dislikes disappointments, on the other hand, may want to. Such preferences correspond (maintaining linearity) to $s = -\delta s'$, $0 < s' < 1$, so that the last two terms in (5) become $\delta^2 E_0^i\left[U_2^i - s' E_1^i\left[U_2^i\right]\right]$. By focussing on $s \geq 0$, I am implicitly assuming that this disappointment-aversion motive, if present, is dominated by anticipatory savoring. All the paper's results can be transposed to the case $s < 0$.

[16] It is straightforward to allow for naiveté, parametrized for instance by a coefficient $\chi \leq 1$ multiplying $(1-q)(1-\lambda^i)$ in (7). This leaves all the positive results unchanged but can affect the welfare conclusions. See Bénabou and Tirole (2002) for such a treatment in a single-agent context.

[17] With imperfect recall, each individual's problem is itself a game of strategic information transmission between his date-0 and date-1 "selves", so there could in general be multiple intrapersonal equilibria. Condition (4) will rule out such multiplicity, which both simplifies the analysis and makes clear that the groupthink phenomenon is really one of collectively rather than individually sustained cognitions.

1.2 Best-response awareness

Following bad news, agents who remain aware that $\theta = \theta_L$ do not exert effort, while those who managed to ignore the news have posterior $r(\lambda^j) \geq q$ and choose $e^j = 1$. Responding as a "realist" to a signal $\sigma = L$ thus leads for agent i to intertemporal expected utility (R is for "realism")

$$U_{0,R}^i = \delta(\delta + s) \left[\alpha \cdot 0 + (1 - \alpha)(1 - \lambda^{-i})\theta_L \right], \tag{8}$$

reflecting his knowledge that only the fraction $1 - \lambda^{-i}$ of other agents who are in denial will exert effort. If he censors, on the other hand, he will assign probabilities $r(\lambda^i)$ to the state being H, in which case everyone exerts effort with productivity θ_H, and $1 - r(\lambda^i)$ to it being really L, in which case only the other "optimists" like him are working and their output is $(1 - \lambda^{-i})\theta_L$. Hence ($D$ is for "denial"):

$$U_{0,D}^i = -m + \delta \left(-c + \delta \left[\alpha + (1 - \alpha)(1 - \lambda^{-i}) \right] \theta_L \right)$$
$$+ \delta s \left(r(\lambda^i)\theta_H + \left(1 - r(\lambda^i) \right) \left[\alpha + (1 - \alpha)(1 - \lambda^{-i}) \right] \theta_L \right). \tag{9}$$

Agent i's incentive to deny reality, given that a fraction $1 - \lambda^{-i}$ of others are doing so, is thus:

$$\left(U_{0,D}^i - U_{0,R}^i \right) /\delta = -m/\delta - [c - (\delta + s)\,\alpha\theta_L] + sr(\lambda^i) \left[(1 - \alpha)\lambda^{-i}\theta_L + \Delta\theta \right]. \tag{10}$$

The second term is the net loss from mistakenly choosing $e^i = 1$ due to overoptimistic beliefs. The third one is the gain in anticipatory utility, proportional to the post-denial belief $r(\lambda^i)$ that the state is H and comprising two effects. First, the agent raises his estimate of the fraction of others choosing $e = 1$, from $1 - \lambda^{-i}$ to 1; at the true productivity θ_L, this contributes $(1 - \alpha)\lambda^{-i}\theta_L$ to his expected welfare. Second, he believes the project's value to be θ_H rather than θ_L, so that when everyone chooses $e = 1$ his welfare is higher by $\Delta\theta = \theta_H - \theta_L$.

The incentive for denial is increasing in the agent's own "habitual" truthfulness λ^i, ensuring a unique fixed point (personal equilibrium). This best response to how others think is characterized by the following properties, illustrated in Figure 1 by the dotted curves.

Proposition 1 (Optimal awareness and the MAD principle). *For any cognitive strategy λ^{-i} used by other agents, there is a unique optimal awareness rate λ^i for agent i, with:*
i) $\lambda^i = 1$ for s up to a lower threshold $\underline{s}(\lambda^{-i}) > 0$, λ^i strictly decreasing in s between $\underline{s}(\lambda^{-i})$ and an upper threshold $\bar{s}(\lambda^{-i}) > \underline{s}(\lambda^{-i})$, and $\lambda^i = 0$ for s above $\bar{s}(\lambda^{-i})$.
ii) λ^i decreases with others' awareness rate λ^{-i} if $\theta_L > 0$, and increases with it if $\theta_L < 0$.
iii) λ^i increases with the degree of spillovers $1 - \alpha$ if $\theta_L > 0$, and decreases with it if $\theta_L < 0$.

The first result is straightforward: the more important anticipatory feelings –the consumption value of beliefs– are to an agent's welfare, the more bad news will be repressed.
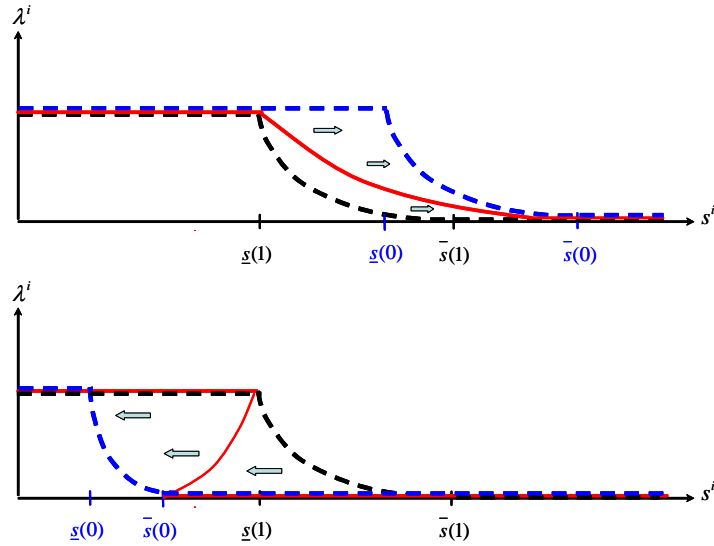
Figure 1: Group Morale ($\theta_L > 0$, upper panel) and Groupthink ($\theta_L < 0$, lower panel). The dotted lines give agent i's optimal awareness λ^i when others are realists ($\lambda^j = 1$) or deniers ($\lambda^j = 0$), with the arrows indicating the transition between the two. The solid lines define the social equilibria.

The second result brings to light a general insight which I shall term the *"Mutually Assured Delusion" (MAD)* principle. If others' wishful thinking leads them to act in a way that is better for an agent than if they were well informed ($\theta_L > 0$), it makes those news not as bad, thus reducing his incentive to engage in denial. But if their avoidance of reality makes things worse than if they reacted appropriately to the true state of affairs ($\theta_L < 0$), future prospects become even more ominous, increasing the incentive to look the other way.[18] In the first case, individual cognitive strategies are strategic *substitutes,* in the latter they are strategic *complements.*

It is worth emphasizing that:

(a) This "psychological multiplier", less than 1 in the first case and greater in the second, arises even though agents' payoffs are completely separable and there is no scope for social learning. It thus represents a novel mechanism giving rise to interdependent beliefs and actions.

(b) The case in which individuals' willful blindness feeds on itself is also that in which it is worse for everyone, as it leads to the wrong course of action ($e^j = 1$ when $\sigma = L$).

• *Public goods and low-risk projects.* The first scenario, best epitomized by a sports team, is that in which an individual's motivation and optimism about the extent to which he can "make a difference" is always valuable to others: effort and quality control in teamwork, recycling, voting, and other forms of good citizenship. More generally, it arises in activities with a limited

---

[18]This argument is for given costs of belief distortion, which is the case here: see (10). An isomorphic one applies when other agents' degree of awareness affects the cost side rather than (or in addition to) the benefit side of an individual's belief manipulations.

downside, in the sense that pursuing them remains socially desirable for the organization even in the low state where the private return falls short of the cost. For a financial institution, one can think of making relatively standard home loans or financing secure brick-and mortar companies, which remains generally profitable even in a mild recession (though less than in a boom).

• *High-risk projects.* The second scenario corresponds to high-risk ventures in which the downside is so bad that persisting in them has negative social value for the group. The archetype is a firm such as Enron, whose strategy is potentially extremely profitable for those involved but may also be completely wrong headed and even illegal, in which case everyone will ultimately suffer heavy losses: loss of job or pension, bankruptcy, even prosecution. Other recent examples include banks investing in dot.com startups, subprime mortgages, CDO's, and the like. The greater is other divisions' or coworkers' tendency –especially among higher-ups, as will be seen below– to ignore red flags and forge ahead with the plan (e.g., set up yet more off-the-books partnerships and other questionable deals or loans), the more catastrophic the losses to be expected if the scheme was flawed, fraudulent, or resting on a bubble. Therefore, the greater the temptation for each employee whose future welfare is tied to the firm's fate to also look the other way, engage in rationalization, and "not think about it".

The proposition's third result shows how both types of cognitive interdependencies are *amplified,* the more closely tied an individual's welfare is to the actions of others.[19] Three interesting implications ensue:

(a) Groupthink phenomena are likely to be particularly important for closed, cohesive groups whose members perceive that they largely share *a common fate* and have few exit options. This is in line with Janis' (1972) findings, but with a slightly different and more precise notion of "cohesiveness".

(b) In groups with asymmetric roles, such as *hierarchies,* there will be a tendency to "follow the leader" into realism or denial. This idea is formalized in Section 1.4 below.

(c) Contagious beliefs are also more likely for large-scale *public goods,* such as those provided by a government, market, or other society-wide institutions which a single individual has little power to affect. This point is pursued in Bénabou (2008), which deals with society-wide ideologies concerning the relative efficacy of markets and governments.

1.3   Social cognition

I now solve for a full social equilibrium in cognitive strategies, looking for fixed points of the mapping $\lambda^{-i} \to \lambda^i$. The main intuition stems from Proposition 1 and is illustrated by the solid lines in Figure 1. First, $\lambda = 1$ is an equilibrium for $s \leq \underline{s}(1)$, as realism is the best response to realism; similarly, $\lambda = 0$ is an equilibrium for $s \geq \bar{s}(0)$, where denial is the best response to

---

[19]This intuition is reflected in (10) through the term $(1 - \alpha) \lambda^{-i} \theta_L$. A lower α also increases the cost of suboptimal effort when $\theta_L > 0$ and raises it when $\theta_L < 0$, reinforcing this effect (term $c - \alpha (\delta + s) \alpha \theta_L$).

denial. Second, when $\theta_L > 0$ (cognitive substitutes), the thresholds $\underline{s}$ and $\bar{s}$ are both decreasing in λ^{-i}, so $\underline{s}(1) < \bar{s}(1) < \bar{s}(0)$ and the two pure equilibria must correspond to distinct ranges. When $\theta_L < 0$ (cognitive complements), on the other hand, both thresholds are increasing in λ^{-i}, and if that effect is strong enough one can have $\bar{s}(0) < \underline{s}(1)$, creating a range of overlap.

Proposition 2 (Groupthink) *1) If the following condition holds,*

$$(11) \qquad\qquad (1 - q)\,(\theta_H - \theta_L) < (1 - \alpha)\,(-\theta_L),$$

then $\bar{s}(0) < \underline{s}(1)$ and for any s in this range, both realism $(\lambda = 1)$ and collective denial $(\lambda = 0)$ are equilibria, with a mixed-strategy equilibrium in between. Under denial agents always choose $e^j = 1$, even when it is counterproductive.

2) If (11) is reversed, $\underline{s}(1) < \bar{s}(0)$ and the unique equilibrium is $\lambda = 1$ to the left of $(\bar{s}(1),\ \underline{s}(0))$, a declining function $\lambda(s) \in (0, 1)$ inside the range, and $\lambda = 0$ to the right of it.

Equation (11) reflects the MAD principle at work. The left-hand side is the basic incentive to think that actions are highly productive (θ_H rather than θ_L) when there are no spillovers ($\alpha = 1$) or, equivalently, when fixing everyone else's behavior at $e = 1$ in both states. The right-hand corresponds to the expected losses –relative to what the correct course of action would yield– inflicted on an individual by others' delusions, and which he can (temporarily) avoid recognizing by denying the occurrence of the bad state altogether. These losses must be of sufficient importance relative to the first, unconditional, motive for denial.

Comparative statics. The proposition also yields a number of potentially testable predictions. First, there is the reversal in how agents respond to others' beliefs (or actions) depending on the sign of θ_L, with the very different equilibrium patterns that result. Second, and focusing on the more interesting case where (11) holds, the expressions for $\bar{s}(0)$ and $\underline{s}(1)$, given in appendix, show that:

a) A more "common fate" of agents (lower α) makes collective denial of the bad state more likely, by lowering both thresholds.

b) A more desirable or more plausible high state (higher θ_H or q) has the same effects.

c) A worse low state (lower $\theta_L < 0$), arising for instance from a more risky project, has more subtle effects. On the one hand, it makes a realistic equilibrium easier to sustain ($\underline{s}(1)$ increases): the cost of making the wrong decision rises, while there is no harmful delusion of others to "escape from". When others are in denial, on the other hand, a lower θ_L makes it even worse. If $1/\alpha - 1/q$ (which must be positive by (11)) is relatively small, the first effect dominates and $\bar{s}(0)$ increases: sufficiently bad news will lead people to "snap out" of their collective delusion. With a sufficiently "common fate" or high priors ($1/\alpha - 1/q$ large enough), on the other hand, the second effect dominates and $\bar{s}(0)$ decreases. The range over which multiplicity occurs thus

widens, and a worsening of the bad news can now cause a previously realistic group to take refuge into groupthink.

The types of enterprises that are most prone to collective delusions are thus: (i) those involving new technologies, products, markets or policies that combine a highly attractive upside and a disastrous downside (representing a mean-preserving spread relative to previous conditions); (ii) those in which participants have only limited exit options and, consequently, a lot riding on the soundness or folly of other's judgements. Such dependence typically arises from irreversible or illiquid prior investments: specific human capital, professional reputation or network, company pension plan, etc. Alternatively, it could reflect the large-scale nature of the problem (e.g., state of the economy, quality of the government, global warming), from which it is hard for a single individual to escape.

1.4 Asymmetric roles: hierarchies and corporate culture

"And if the blind lead the blind, both shall fall into the ditch." (Matthew 15:14)

I now demonstrate the generality of the MAD principle by relaxing all the symmetry assumptions, as well as the state-invariance of the payoff to "inaction" ($e = 0$). I then use this more general result to show how, in hierarchical organizations, denial and realism will "trickle down". Let the payoff structure (1) be extended to:

$$(12) \qquad U_2^i \equiv \sum_{j=1}^n \left(a_\sigma^{ji} \, e^j + b_\sigma^{ji} \, (1 - e^j) \right), \quad \text{for all } i = 1, \ldots n \text{ and } \sigma \in \{H, L\}.$$

Each agent j's choice of $e^j = 1$ thus creates a state-dependent value a_σ^{ji} for agent i, while $e^j = 0$ generates value b_σ^{ji}; for $i = j$, these correspond to agent i's private returns to action and inaction. All payoffs remain linearly separable for the same expositional reason as before, but complementarities or substitutabilities can easily be incorporated, as shown in Section 1.6 below. Agents may also differ in their preference and cognitive parameters c^i, m^i, δ^i, in their proclivity to anticipatory feelings s^i or even in their priors q^i. The generalization of (4) is thus

$$(13) \qquad a_L^{ii} - b_L^{ii} < \frac{c^i}{s^i + \delta^i} < q^i \left(a_H^{ii} - b_H^{ii} \right) + (1 - q^i) \left(a_L^{ii} - b_L^{ii} \right),$$

while the generalization of $\theta_H > \theta_L$ (H is the better state, conditional on everyone taking the optimal action), is

$$(14) \qquad \sum_{j=1}^n a_H^{ji} > \sum_{j=1}^n b_L^{ji}.$$

Focussing here on pure-strategy equilibria, one can again compare an agent i's incentive to ignore a signal $\sigma = L$ when surrounded by deniers ($\lambda^j \equiv 0$) and by realists ($\lambda^j \equiv 1$). The condition for

11

complementarity, generalizing $\theta_L < 0$, is now:

$$(15) \qquad \sum_{j \neq i} \left(a_L^{ji} - b_L^{ji} \right) < 0, \quad \text{for all } i = 1, \ldots n.$$

In accordance with the MAD principle, it means that other's delusions, leading them to choose $e^j = 1$ even when $\sigma = L$, are *on average* harmful to agent i. Multiple equilibria occur when this expected loss is sufficiently large relative to the "unconditional" incentive to deny:

$$(16) \qquad (1-q) \sum_{j=1}^{n} \left(a_H^{ji} - a_L^{ji} \right) < \sum_{j \neq i} \left(b_L^{ji} - a_L^{ji} \right).$$

Proposition 3 (Organizational cultures) *Let (13), (14) and (16) hold for all $i = 1, \ldots n$. There exists a non-empty range $\left[\bar{s}^i(0), \underline{s}^i(1) \right]$ for each i, such that if $(s^1, \ldots s^n) \in \Pi_{i=1}^{n} \left[\bar{s}^i(0), \underline{s}^i(1) \right]$, then both collective realism ($\lambda^i \equiv 1$) and collective denial ($\lambda^i \equiv 0$) are equilibria.*

• *Directions of cognitive influence* Going beyond multiplicity, interesting new results emerge for organizations in which members play asymmetric roles. Indeed, the thresholds $\bar{s}^i(0)$ and $\underline{s}^i(1)$, given in the appendix, confirm the intuition that each agent's optimal awareness is most sensitive to how the people whose decisions have the greatest impact on his welfare (the largest contributors to (15)) deal with unwelcome news. As an application, consider the simplest form of hierarchy: two agents, 1 and 2, such as a manager and worker. If $a_L^{12} - b_L^{12}$ is sufficiently negative while $\left| a_L^{21} - b_L^{21} \right|$ is relatively small, agent 2 suffers a lot when agent 1 loses touch with reality, while the converse is not true.[20] Workers thus risk losing their job if management makes overoptimistic investment decisions, whereas the latter has little to lose (perhaps the reverse) if workers put in more effort than realistically warranted. When the asymmetry is sufficiently pronounced (conditions are given in the appendix), this leads to a (testable) pattern of predominantly *top-down cognitive influences,* illustrated in Figure 2. Formally,

$$(17) \qquad \left[\underline{s}^1(1), \bar{s}^1(0) \right] \subset \left[\bar{s}^2(0), \underline{s}^2(1) \right] \equiv S$$

and for all $(s^1, s^2) \in S \times S$ there is a *unique equilibrium,* such that:

(a) The qualitative nature of agent 1's cognitive strategy –complete realism, complete denial, or mixing– depends only on s^1, not on s^2.

(b) If agent 1 behaves as a systematic denier (respectively, realist), so does agent 2: where $\lambda^1 = 1$ it must be that $\lambda^2 = 1$, and similarly $\lambda^1 = 0$ implies $\lambda^2 = 0$.

c) Only when both agents are in partial denial (between two loci in Figure 2) does agent 2's degree of realism influence that of agent 1.

---

[20] Consequently, agent 2's cognitive strategy will have strong positive dependence on that of agent 1, $(\bar{s}^2(0) < \underline{s}^2(1)$, as in the bottom panel of Figure 1), while that of agent 1 will vary little with that of agent 2 $(\underline{s}^1(1) < \bar{s}^1(0))$.
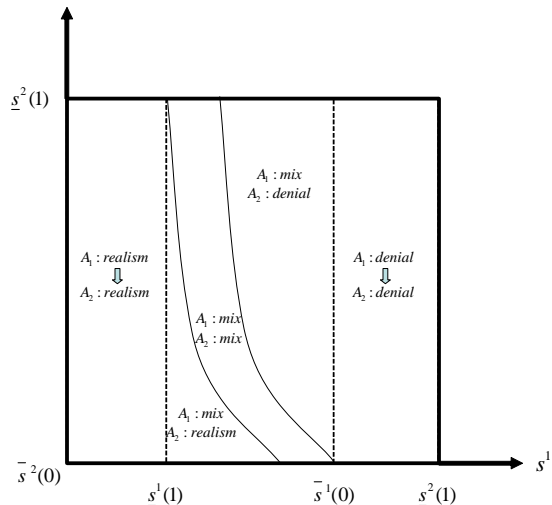
Figure 2: "Trickle down" of realism and denial in a hierarchy

Let agent 2 now be replicated into $n-1$ identical "workers", each with influence $[a_\sigma^{j1} e^j + b_\sigma^{j1} (1 - e^j)]/(n-1)$ over the manager or "leader", but subject to the same influence from him as before, $a_\sigma^{1j} e^1 + b_\sigma^{1j} (1 - e^1)$. Figure 2 then remains operative, showing how *the leader's attitude toward reality* will tend to *spread to all his subordinates,* while being influenced by theirs only in a limited way, and over a limited range.

This result has clear applications to corporate and bureaucratic culture, explaining how people will contagiously invest excessive *faith in a leader's "vision"*.[21] It also has relevance to the political sphere. Thus, a dictator who is secure in his power need not exert constant censorship or constraint to implement his policies, as crazy as they may be: he can rely on people's mutually reinforcing tendencies to rationalize as "not so bad" the regime they (endogenously) have to live with.

The present model is of course an oversimplified representation of an organization; yet the same general principles should carry over to more realistic hierarchies, with multiple tiers, strategic interactions, control rights, transfer payments, losers and gainers from the delusions of others, etc. I leave such extensions to future work, and return from here on to the basic, symmetric model of Section 1.1.

---

[21] In Rotemberg and Saloner (1993), a manager's "vision" (prior beliefs or preferences that favor some types of activities over others) serves as a commitment device to reduce workers' concerns about ex-post expropriation of their innovations. In Prendergast (1993), the use by managers of subjective performance evaluations to assess subordinates' effort at seeking new information leads workers to distort their reports in the direction of the manager's (expected) signal. Both mechanisms thus lead workers to "conform" their behavior to managers' prior beliefs; unlike here, however, in neither case do they actually espouse those beliefs, nor would the manager ever want them to report anything but the truth.

1.5 Welfare, shooting the messenger, and free-speech guarantees

Are agents in collective denial worse or better off than if they squarely faced the truth –as an alternative equilibrium, or possibly by means of some collective commitment mechanism? Conversely, can they benefit from preserving a high morale if everyone is able commit to ignoring bad news?

Consider first state $\sigma = L$, which occurs with probability $1 - q$. When agents are realists (setting $\lambda^j = 1$ in (8)), equilibrium welfare is $U_{L,R}^* = 0$. When they are deniers (setting $\lambda^j = 0$ in (9)), it is given by:

$$(18) \qquad U_{L,D}^*/\delta = -m/\delta - c + \delta\theta_L + sq\theta_H + s(1-q)\theta_L.$$

Collective denial following bad news is thus harmful or beneficial, depending on whether s is below or above the threshold

$$(19) \qquad s^* \equiv \frac{m/\delta + c - \delta\theta_L}{q\theta_H + (1-q)\theta_L},$$

as illustrated in Figure 3.

Proposition 4 *Welfare following bad news (state L):*
1) If $\theta_L < 0$, then $s^* > \max\{\bar{s}(0), \underline{s}(1)\}$, so whenever realism ($\lambda = 1$) is in the equilibrium set, it is superior to denial ($\lambda = 0$). Moreover, there exists a range in which realism is not an equilibrium but, if it can be achieved through collective commitment, leads to higher welfare.
2) If $\theta_L > 0$, then $s^* < \bar{s}(0)$. The equilibrium thus involves excessive realism for $s \in (s^*, \bar{s}(0))$ and excessive denial for $s \in (\underline{s}(1), s^*)$, when this interval is nonempty.

Given how damaging collective delusion is in state L with $\theta_L < 0$, it makes sense that when realism can also be sustained as an equilibrium it dominates, and that when it cannot the group may try to commit to it. Conversely, with $\theta_L > 0$, *boosting morale* in state L is *helpful* to overcome the free-rider problem, so the group would want to commit to ignoring bad signals when $s \geq s^*$ but the only equilibrium involves realism.[22]

Consider now welfare in state H, which occurs with probability q: given (4), everyone chooses $e^i = 1$ in both equilibria. Under denial, however, agents are unsure of whether the state is truly H, or it was really L and they censored the bad news. As a result of this "spoiling" effect, welfare is only

$$(20) \qquad U_{H,D}^*/\delta = -c + \delta\theta_H + s[q\theta_H + (1-q)\theta_L] < -c + (\delta + s)\theta_H = U_{H,R}^*/\delta.$$

---

[22] If θ_L is high enough that $\delta\theta_L > c + m/\delta$, then $s^* < 0$. Denial in state L is then socially beneficial even absent anticipatory emotions ($s = 0$): high group morale helps alleviate the free-rider problem, e.g., in a sports team.
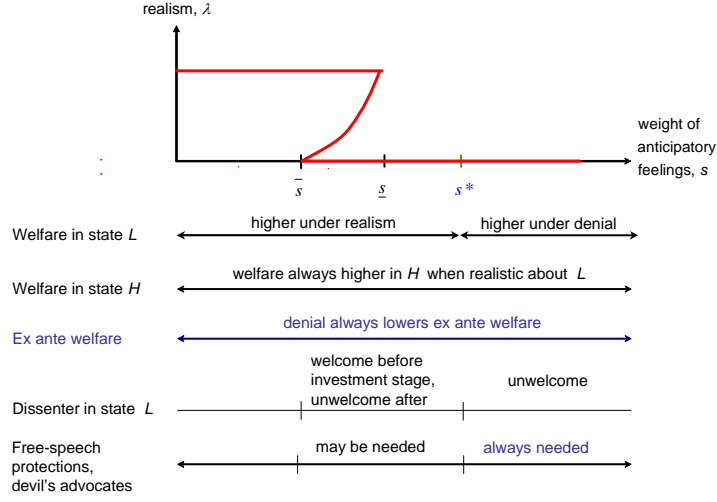
Figure 3: Welfare and dissenting speech (groupthink case)

Averaging over the two states, finally, the mean belief about θ remains fixed (by Bayes' rule), so the net welfare impact of denial is just

$$(21) \qquad \Delta W \equiv (1 - q) \left[(\delta + s) \, \theta_L - c - m/\delta \right],$$

realized in state L. In assessing the overall value of social beliefs, one can thus focus only on *material* outcomes and ignore anticipatory feelings, which are much more difficult to measure but average out across states of nature.[23]

Proposition 5 *1) Welfare following good news (state H) is always higher, the more realistic agents are when faced with bad news (the higher is λ).*

(2) If $\theta_L \leq 0$, denial always lowers ex-ante welfare. If $\theta_L > 0$, it improves it only for $(\delta + s) \, \theta_L > c + m/\delta$.

These results, also illustrated in Figure 3, lead to a clear (and potentially testable) distinction between two types of collective beliefs and the situations that give rise to them.

• *Beneficial group morale.* When $\theta_L > 0$, $e = 1$ is socially optimal even in state L, but since $\alpha(s + \delta)\theta_L < c$ it is not privately optimal. If agents can all manage to ignore bad news at relatively low cost, either as an equilibrium or through commitment, they will thus be better off both not only *ex-post* but also *ex-ante:* $\Delta W > 0$ This is in line with a number of recent results showing the functional benefits of overoptimism (achieved through information manipulation or

---

[23] As long as agents are Bayesian, which seems like a reasonable assumption for types of activities in which they engage recurrently.

appropriate selection of agents by a principal) in settings where agents with the correct beliefs would underprovide effort.[24]

• *Harmful groupthink.* The novel case is the one in which contagious delusions can arise, $\theta_L < 0$, and it also leads to a more striking conclusion: not only can such reality avoidance greatly damage welfare in state L, but even when it improves it those gains are always dominated by the losses induced in state H, so $\Delta W < 0$.[25] This normative result also has positive implications for how organizations deal with dissenters, revealing an interesting form of time inconsistency between *ex ante* and *ex post* attitudes. In carrying out this discussion, I shall refer interchangeably to "the group" and to "society", as in the case of political ideologies.

The curse of Cassandra. Let $\theta_L < 0$ (more generally, $(\delta + s)\theta_L < c + m/\delta$) and consider a denial equilibrium, as illustrated in Figure 3. Suppose now that, in state L, an individual or subgroup with a lower s or a different payoff structure attempts to bring back the facts to everyone's attention. If this occurs after agents have have sunk in their investment it simply amounts to deflating expectations in (3), so they will refuse to pay attention, or may even try to "kill the messenger" (pay a new cost to forget). Anticipating that others will behave in this way, in turn, allows everyone to more confidently invest in denial at $t = 0$. To avoid this deleterious outcome, organizations and societies will find it desirable to set up *ex-ante* guarantees such as whistle-blower protections, devil's advocates positions, constitutional rights to free speech, independence of the press, etc. These will ensure that bad news will most likely "resurface" *ex-post* in a way that is hard to ignore, thus lowering the return (or raising the cost) of investing in denial.

Similar results apply if the dissenter brings his message at an interim stage, after people have censored but before investments have been made. For $s < s^*$ they should, in principle, welcome the opportunity to collectively return to reality and correct course. In practice, this may be hard to achieve: it may not be an equilibrium (case $\theta_L > 0$), or require full coordination (case $\theta_L < 0$). With payoff heterogeneity, dissenters' motives may also be suspect, making it hard to convince others. The conclusion is even starker if people value maintaining hope (or dislike anxiety) sufficiently that $s > s^*$. In that case, bringing (back) the bad news about the state really being L will hurt everyone, leading to a universal unwillingness to listen and rejection –the curse of Cassandra. And yet, free-speech guarantees and mechanisms encouraging dissent

---

[24]In a team or firm context see, e.g., Bénabou and Tirole (2003), Gervais and Goldtstein (2005), Fang and Moscarini (2005) and Van den Steen (2005). In a self-control context, see Carrillo and Mariotti (2000), Bénabou and Tirole (2002) and Battaglini et al. (2005). Also closely related to the present framework is Dessi (2005), who shows how one generation may want to collude in order to paint to the next one an overly optimistic picture of the benefits of cooperation. Dessi studies only the social-planner solution achieved through centralized control of beliefs (e.g., by an all-powerful state), and thus does not consider what equilibria arise from parents' individual child-rearing and indoctrination decisions, or their own ideological choices.

[25]The "shadow of doubt" cast over the good state by the censoring of the bad state could also distort some decisions in state H, although in this simple example it does not. Conversely, departing from Bayesian updating, for instance by introducing in (7) a "naivete" coefficient $\chi \leq 1$ multiplying $1 - \lambda$, would attenuate the losses in state H and thus allow ex-ante gains. See Bénabou and Tirole (2002, 2006c) for examples of both effects.

remain desirable ex-ante, because they avoid welfare losses in state H and, on average, save the organization or society from wasting resources on denial (including killing messengers). There is now a strong tension between *ex-ante* and *ex-post* incentives to tolerate dissenting speech, illustrations of which abound in corporations, bureaucracies, and polities.

1.6   Strategic interactions

To highlight the model's new source of interdependence in beliefs and behaviors, I have until now focussed attention on standard public-goods settings in which an agent's welfare level depends on others' actions, but his return to acting does not. The presence of strategic complementarities will, quite intuitively, reinforce the tendency for contagion, whereas substitutabilities will work against it.[26] To see this, let agent i' expected payoff in state $\sigma \in \{H, L\}$ now be $\Pi_\sigma(e^i, \mathbf{e}^{-i})$, where $\mathbf{e}^{-i}$ denotes the vector of others' actions; his incentive to act is then $\pi_\sigma(\mathbf{e}^{-i}) \equiv \Pi_\sigma(1, \mathbf{e}^{-i}) - \Pi_\sigma(0, \mathbf{e}^{-i})$.

In state L, the differential in i's anticipatory value of denial that results from others' "blind" persistence, previously given by $-s(1-\alpha)\theta_L$, is now

$$-s \sum_{j \neq i} \left[\Pi_L(1, \mathbf{0}) - \Pi_L(1, \mathbf{1}) \right],$$

which embodies the exact same (MAD) intuition as before. The new ingredient is that others' persistence now also changes the material value of investing in state L (previously a fixed $\alpha\theta_L$), by an amount equal to

$$\sum_{j \neq i} \left[\pi_L(\mathbf{1}) - \pi_L(\mathbf{0}) \right],$$

with sign governed by $\Sigma_{j \neq i} \, \partial^2 \Pi_L(e^i, \mathbf{e}^{-i}) / \partial e^i \partial e^j$. When actions are complements, delusion is thus less costly if others are also in denial, whereas with substitutes it is more costly. Rather than restate general results with nonseparable payoffs, which is not hard but would not add much insight, I shall focus on an important concrete application: how, in spite of output decisions being substitutes, asset markets can be seized by collective "manias", ultimately followed by a crash.

2   "Irrational" market exuberance

"For Countrywide, the quintessential proxy for the mortgage meltdown that now surrounds it, this remains one of the burning and still unanswered questions. Why did the company's chief, who routinely warned of his rivals' lax lending practices well before the mortgage market

---

[26] At the same time, without anticipatory feelings (or some other "non-standard" role for beliefs), no amount of complementarity can generate results similar to those of the model: agents with standard preferences always have (weakly) positive demand for information, and thus never engage in denial or ex-post rationalizations.
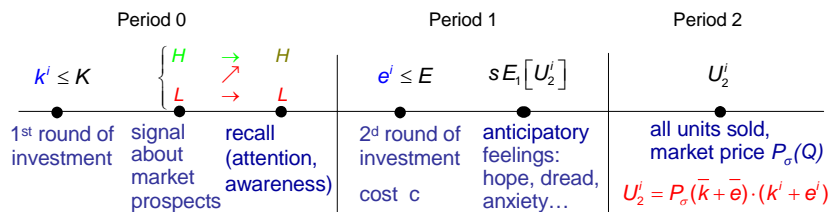
Figure 4: The market game

cracked, ultimately allow Countrywide to ardently embrace those practices?... According to... a former banking analyst and founder of a New York investment fund, 'The biggest self-inflicted wound here is they should have pulled back in '05 and '06 when you had these competitors doing all sorts of crazy things. Angelo [Mozilo] talked about the danger but somehow went for the market share gains anyway.'" (Morgenson and Fabrikant (2007))

I consider now a continuum of firms or investors operating in a market with the following "time to build" features. Each agent i can produce $k^i \leq K$ units of a good (housing, office space, internet startup firm) in period 0 and an additional $e^i \leq E$ units in period 1, where K and E reflect capacity constraints or similar technological limits. The cost of production in period 0 is set to 0 for simplicity, while in period 1 it is equal to c. All units are to be sold at $t = 2$, at which time the *expected* market price $P_\sigma(\bar{k} + \bar{e})$ will reflect total supply, $\bar{k} + \bar{e} \in [0, K + E]$, as well as stochastic market conditions, θ_σ, with $\sigma = H, L$. In-between the two production phases, agents all observe the signal σ, then decide how to process it, with the same information structure and preferences as before. See Figure 4.

To take recent examples, θ_H may correspond to a "new economy" in which high-tech startups will flourish and their prospects are best assessed using "new metrics"; to a permanent rise in housing values; or to any other positive and lasting shift in fundamentals. Conversely, θ_L would reflect an inevitable return to "old" economy and valuations, the presence of a bubble that will ultimately burst, or the unsustainability for many households of meeting future payments on their adjustable-rate mortgages, stated-income loans and other subprime debt. Finding reasons to believe in H even as evidence of L accumulates then corresponds to what Shiller (2005) terms "new-era thinking", and of which he provides many examples.

The absence of an interim or futures asset market for the good before date 2 is a version (chosen for simplicity) of the kind of "limits to arbitrage" commonly found in the finance literature. Specifically, I assume that: (i) goods produced in period 0 cannot be sold before period 2, for instance because they are still work-in-progress whose quality or market potential is not verifiable: startup company, unfinished residential development or office complex, new type of financial asset; (ii) short sales are not feasible.

Empirically, such limited arbitrage possibilities seem quite descriptive of the types of markets

18

	Lehman Brothers	Bear Stearns	
Level 1	96 (35.6%)	39 (17.7%)	Trade in active markets with readily available prices
Level 2	152 (56.3%)	163 (74.1%)	"Mark to model"
Level 3	22 (8.2%)	18 (8.2%)	"Reflect management's best estimates of what market participants would use in pricing the assets"
Total ($ billions)	270	220	

Figure 5: Financial assets on balance sheet, 2d fiscal quarter of 2007. Source: Reilly (2007).

which the model aims to analyze.[27] In the recent mortgage-related crisis, for instance, a dominant fraction of the assets held by major U.S. investment banks did not have an active trading market in which they could be objectively priced. Instead, they were valued according to the bank's own model and projections, or even according to management's "best estimates". Figure 5 shows the figures for Lehman Brothers and Bear Stearns, constructed from Reilly (2007).[28] In housing, similarly, regional-index futures (Case-Shiller) are a very recent innovation and their market is still small and fairly illiquid.

I shall assume (and later provide conditions ensuring) that, ex-ante, the market is sufficiently profitable that everyone will invest to full capacity in period $0: k^j = \bar{k} = K$. Moreover, following (4), let

$$(22) \qquad P_L(K) < \frac{c}{s+\delta} < \frac{c}{\delta} < qP_H(K+E) + (1-q)P_L(K+E).$$

Thus, it is a dominant strategy for a firm at $t = 1$ to produce the maximum $e^i = E$ if its posterior is no worse than its prior q, and not to produce if it knows for sure that the state is L.

I now analyze the market subgame that unfolds when agents observe the signal L. The optimality of first-stage investment $k^j = K$ (which involves expected profits in both states) is shown in the appendix and taken here as given, for expositional simplicity.

　　1. *Realism.* Suppose first that market participants acknowledge and properly respond to

---

[27]Schiller (2003) discusses the costs and risks of short-selling He also cites studies documenting the fact that in recent times, short sales never amounted to more than 2% of stocks (whether in number of shares or value). Gabaix et al. (2007) provide recent econometric evidence of limits to arbitrage in the market for mortgage-backed securities.

[28]The share of Level 3 assets, whose valuations Reilly describes as "little more than management's guesses", was as high as 10% when Goldman Sachs and J.P. Morgan were included, and around 6% when Merrill Lynch was added. Reilly also reports that prior to 2007, accounting rules did not even require firms to break down their assets into these three valuation categories, and that even during the first half of the year "investors didn't pay much attention to this new data". In an interesting sign of wishful groupthink , 30% of Bear Stearns' stock was owned, until the end, by its employees.

bad news: $\lambda^j \equiv 1$. They will then not produce any additional units at $t = 1$, so the price at $t = 2$ will be $P_L(K)$. For an individual investor i with stock k^i, the net effect of ignoring the signal is thus

$$
(23) \qquad
\begin{aligned}
(U^i_{0,D} - U^i_{0,R})/\delta \;=\; & -m/\delta + [(\delta + s)P_L(K) - c]\,E \\
& + sr(\lambda^i)\,[P_H(K + E) - P_L(K)]\,(k^i + E).
\end{aligned}
$$

The second term reflects the expected losses from producing at $t = 1$, while the last one represents the value of maintaining hope that the market is strong or will eventually recover, in which case total output will be $K + E$ and the price $P_H(K + E)$. Realism is an equilibrium if $U^i_{0,D} \leq U^i_{0,R}$ for $\lambda^i = 1$ and $k^i = K$, or

$$
(24) \qquad
s \leq \frac{m/\delta + [c - \delta P_L(K)]\,E}{[P_H(K + E) - P_L(K)]\,(K + E) + P_L(K)E} \equiv \underline{s}(1).
$$

2. Denial. If all other participants remain bullish in spite of adverse signals, they will keep producing at $t = 1$, *causing the already weak market to crash:* at $t = 2$, the price will fall to $P_L(K + E) < P_L(K)$. The net value of denial for agent i is now

$$
(25) \qquad
\begin{aligned}
(U^i_{0,D} - U^i_{0,R})/\delta \;=\; & -m/\delta + [(\delta + s)P_L(K + E) - c]\,E \\
& + sr(\lambda^i)\,[P_H(K + E) - P_L(K + E)]\,(k^i + E).
\end{aligned}
$$

In the second term, the expected losses from overproduction are higher than when other participants are realists. Through this channel, which reflects the *substitutability* of output decisions in a market interaction, each individual's cost of delusion increases when others are deluded. On the other hand, the third term makes clear that the anticipatory value of denial is also greater, since acknowledging the bad state now requires *recognizing an even greater capital loss* on the preexisting holdings. This is again the MAD principle at work.

Denial is an equilibrium if $U^i_{0,D} \geq U^i_{0,R}$ for $\lambda^i = 0$ and $k^i = K$, or

$$
(26) \qquad
s \geq \frac{m/\delta + [c - \delta P_L(K + E)]\,E}{q\,[P_H(K + E) - P_L(K + E)]\,(K + E) + P_L(K + E)E} \equiv \bar{s}(0).
$$

From (24)-(26), it easy to see that bullish denial, followed by market collapse, is more likely (and realism less likely), the greater the accumulated stocks. The model thus generates both:

(a) *Escalating commitment* at the individual level: the more an agent has produced or invested to date (k^i), the more likely he is to continue even in the face of bad news, thus displaying a form of the *sunk cost fallacy*.[29]

---

[29] This effect is closely related to the escalating commitment studied in Bénabou and Tirole (2007), but occurs in that paper through a somewhat different mechanism (self-signaling).

(b) *Market momentum:* the greater the aggregate level of prior production or investment (K), the more likely is each individual is to continue even in the face of bad news.

3. Contagious exuberance. To capture the phenomenon of market manias, new-era thinking, collective blindness to impending crashes, etc., I now examine when other participants' exuberance makes each individual more likely to also be exuberant. Intuitively, this occurs when the substitutability effect, which bears on the *marginal* units E produced in period 1, is dominated by the capital-loss effect on the *outstanding position* K inherited from period 0. Formally, $\bar{s}(0) < \underline{s}(1)$ requires that K be large enough relative to E.

Proposition 6 *(Market manias and crashes) If*

$$(27) \qquad P_H(K+E)\left(\frac{K+E}{K}\right) < \frac{c}{\delta},$$

there exists $q^* < 1$ such that, for all $q \in [q^*, 1]$, there is a non-empty interval for s in which both realism and blind "exuberance" in the face of adverse news are equilibria, provided m is not to large. The latter case leads to overinvestment and eventually to a market crash.

Besides providing an explicit and psychologically based model of investment frenzies and ensuing crashes, the model identifies some key features of the markets that are prone to such cycles. First, there must be a "story" about shifts in fundamentals that is minimally plausible a priori (q must not be too low): technology, demographics, globalization, etc. The key result is then that investors's beliefs in the story can quickly become resistant to nearly all evidence.[30] Moreover, when the new opportunity first appears (q rising above the threshold), there will be an initial phase of investment buildup and rising prices. Finally, the assets in question must be characterized by both significant uncertainty and limited liquidity, as discussed before. These conditions typically apply for assets tied to new technologies, financial instruments or policy regimes (e.g., deregulation), whose potential will take a long time to be fully revealed. They also recur through most of the episodes of stock and housing market frenzies surveyed by Shiller (2005).

There are several ways in which this simple market model could be extended. First, in a dynamic context, outstanding stocks could result (stochastically) from the combination of previous investment decisions and demand realizations. Second, one could relax the relatively strong form of "limits to arbitrage" imposed here through the assumption that trades occur only at $t = 2$ (no forward market). Such "early" trades could instead involve transactions costs, risk due to limited market liquidity or, for large positions, an adverse price impact.[31]

---

[30] This is contrast to traditional models of bubbles arising from an infinite horizon, in which everyone realizes that they are trading a "hot potato" whose value must eventually collapse, and could do so at any instant.

[31] Trying to sell (or sell short) in period 1 could also be self-defeating, as it would reveal (again) to the market that the state is L, generating an immediate price collapse.

• *Direct and indirect stakes.* The analysis leading to Proposition 6 shows how an agent's propensity to respond to danger signals with a "suspension of disbelief" rises with his initial market position or inventory k^i (or the part of it that cannot be easily diversified or quickly unwound), which may differ across participants. Other, more indirect stakes have similar effects, both contributing to and feeding on the propagation of collective exuberance (and ultimate losses) to broad parts of the economy. Indeed, if indicators point to a state of the world in which the high-tech or housing sector is headed for a crash and the economy for a recession, all three major assets of households are at risk: their job, the value of their house and their pension, the latter especially is some of it is invested in their employer's stock. The worse the potential crisis is made by other agents' feeding of the market frenzy, the greater is the incentive not to acknowledge these risks (ignore or rationalize away the signal $\sigma = L$). And, as a result, the greater the likelihood that the household will itself contribute to the excessive buildup of debt, housing, or undiversified stock holdings.

Another set of key actors with "value at risk" are politicians and regulators, whose career and reputation will be badly damaged if the disaster scenario (state L, worsened by market participants' willful blindness) occurs. While this should normally make them try to dampen the market enthusiasm and buildup, it has proceeded far enough (high K) that large, economy-wide losses are are unavoidable in the bad state, they may also become "believers" in a rosy future or smooth landing. Consequently, they will fail to take the measures that could have limited (tough not avoided) the damage, and thus further enable the market mania and subsequent crash.[32] In some cases, public officials may also have an "ideological" stake in (say) the virtues of unfettered financial markets: a severe crisis that would publicly prove such faith to be excessive would reduce the general credibility of laissez-faire arguments and increase demand for public regulation in other parts of the economy.

3 Other applications and extensions

3.1 Collective apathy and fatalism

The form of denial considered so far has been a collective "illusion of control" or overconfidence, leading a group, organization or market to persist in a costly course of action in spite of widely available evidence that it is doomed. The opposite case is collective apathy: rather than ac-

---

[32]From Andrews (2007): "Edward M. Gramlich, a Federal Reserve governor who died in September, warned nearly seven years ago that a fast-growing new breed of lenders was luring many people into risky mortgages they could not afford. But when Mr. Gramlich privately urged Fed examiners to investigate mortgage lenders affiliated with national banks, he was rebuffed by Alan Greenspan, the Fed chairman... Mr. Greenspan and other Fed officials repeatedly dismissed warnings about a speculative bubble in housing prices. In December 2004, the New York Fed issued a report bluntly declaring that "no bubble exists." ... The Fed was hardly alone in not pressing to clean up the mortgage industry. When states like Georgia and North Carolina started to pass tougher laws against abusive lending practices, the Office of the Comptroller of the Currency successfully prohibited them from investigating local subsidiaries of nationally chartered banks".

knowledge a crisis that could be partly remedied through timely action, everyone pretends that things, though perhaps not great, "could be worse", and that little can be done to improve them anyway. One can think of an ethnic group subject to discrimination or threatened by another one, but whose members pessimistically deem it useless to fight back, try to escape or otherwise improve their lot (see, e.g., Cialdini (1984) and Hochschild (1996) on minorities' acquiescence to a discriminatory system). Another example, examined below, is that of "tuning out" the distress of others.

To capture these ideas, I simply extend (1) to

$$(28) \qquad\qquad U_2^i = \theta \left[\alpha e^i + (1 - \alpha)e^{-i} - \kappa \right].$$

• When $\kappa < \min\{1, \theta_H/\Delta\theta\}$, state H remains (conditional on $e = 1$) a more favorable state than L, and one can show that for κ below a certain threshold all the results of the case $\kappa = 0$ carry over with little change. Indeed, $-\kappa > 0$ plays a role similar to the outstanding market positions K in the previous section.

• When $\kappa > \max\{1, \theta_H/\Delta\theta\}$, on the contrary, state H corresponds to a *crisis state:* action is called for but, even when carried out effectively ($e^j \equiv 1$), will not suffice to offset the shock, leaving agents worse off than in state L. Intuition now suggests that an equilibrium in which agents respond appropriately to crises can coexist with one in which they systematically censor such signals and *always* remain passive.[33]

Indeed, this problem is closely related to the original one, once recast in terms of the relative effectiveness of *inaction*. Formally, let $\tilde{\theta}$ take values $\tilde{\theta}_{\tilde{H}} \equiv -\theta_L$ in state $\tilde{H} \equiv L$ and $\tilde{\theta}_{\tilde{L}} \equiv -\theta_H < 0$ in state $\tilde{L} \equiv H$, with respective probabilities $\tilde{q} \equiv 1 - q$ and $1 - \tilde{q}$; similarly, let $\tilde{c} \equiv -c$. Using these transformed variables, it is then easy to obtain "parallels" to Propositions 2 to 5. In particular, condition (4) is replaced by

$$(29) \qquad\qquad q\theta_H + (1 - q)\,\theta_L < \frac{c}{\alpha\,(s + \delta)} < \frac{c}{\alpha\delta} < \theta_H,$$

and the equilibrium strategies and thresholds are obtained by replacing $\Delta\theta$ with $-\kappa\Delta\theta$ and θ_H, θ_L, q, and c with their "tilde" analogues.

Proposition 7 *Assume (29) and $\kappa > \max\{1, \theta_H/\Delta\theta\}$. All the results in Proposition 2 remain, but with denial ($\lambda < 1$) now occurring in state H only and leading to inaction. Multiple equilibria occur if and only if $q\,(\kappa\Delta\theta) < (1 - \alpha)\,\theta_H$.*

The left-hand side of this modified MAD condition reflects the action-independent gain from being in the no-crisis state, while the right-hand side measures the losses inflicted by all those

---

[33]Furthermore, there is now no equilibrium in which agents censor the signal $\sigma = L$, just like when $\kappa = 0$ (or κ sufficiently below $\min\{1, \theta_H/\Delta\theta\}$ more generally) there is no equilibrium in which they censor $\sigma = H$. See Lemma 4 and the proof of Proposition 7 in the appendix, with $\Delta\gamma \equiv -\kappa\Delta\theta$.

who, denying that a crisis has occurred, *fail to act.*

• *Helping others or tuning out.* Studies of how people respond to the distress of others – victims of accidents, wars, natural disasters, famine, etc.– display two important puzzles. First, people show a greater willingness to help or contribute when the number of those perceived to be in need is small than when it is large. Slovic (2007) discusses a number of experiments documenting such "psychic numbing" (lowered affective reactions and willingness to give) in response to even small absolute increases in the size of the at-risk group. He further argues for the importance of this phenomenon in accounting for recurrent public inertia in the face of humanitarian disasters, poverty, mass persecutions, and the like. A second regularity, common to most public-goods situations, is that people give and help more when they know that others are doing so.[34]

The above results can help understand both phenomena. Let K be the number of people in need, or emphasized as being in need, and let θ be the severity of their situation. For a given cost c, each individual $i = 1, \dots n$ can help up to a victims ($e^i = 1$), and he experiences an empathic disutility equal to the total amount of suffering,

$$(30) \qquad U_2^i = -\theta \left[K - a\Sigma_{j=1}^n e^j \right].$$

Note that this *does not assume* that people intrinsically undervalue "statistical lives" or actions that represent only "a drop in the ocean". Instead, this will be a *result*. Indeed, (30) clearly corresponds to the model, with $\alpha = 1/n$, $\kappa = K/na$ and θ simply replaced by θna. Therefore, as K increases beyond a critical threshold:

a) The loss in utility from acknowledging $\theta = \theta_H$ overtakes an individual's ability to remedy it, leading him to switch from helping to "tuning out" the problem altogether. Thus, he effectively censors from awareness and recall all painful evidence of the crisis: turning the page of the newspaper, switching the channel, rationalizing the situation as not so bad, etc.

b) The level at which an individual switches from response to non-response depends on how many others he believes are helping or also tuning out: what matters to i is $K - a\Sigma_{j\neq i}^n e^j$. Hence, within some range of K, both *collective generosity* and *collective apathy* –what Slovic terms the "collapse of compassion"– are social equilibria, even though charitable giving involves (realistically) no increasing returns.

c) Vivid, memorable images of the *intensity* of individual suffering θ (but *not* the number, K, which has the reverse effect) make the crisis more difficult to put "out of mind" and thus

---

[34] An alternative explanation for this second set of findings is one of social or personal norms; see Bénabou and Tirole (2006b). The first phenomenon is distinct from (but often combines with) the "identifiable victim effect". For instance, Small et al. (2007) found that donations to a specifically identified Malawian child facing the risk of starvation decreased considerably when information about the child was complemented with background statistics documenting the scale of food shortages in Africa (donations fell by more than a half, to a level close to that observed when only impersonal statistics were provided).

reduce the scope of apathy. In the multiplicity range, one small such example, widely publicized, may trigger a large equilibrium shift.

3.2 Other informational preferences or technologies

[[To be added]]

4 Conclusion

This paper has developed a model of how wishful thinking and reality denial or rationalizations spread through organizations and markets. The underlying mechanism relies neither on strategic complementarities, agents' herding on a private subset of private signals, or exogenous biases in inference. It is also widely applicable, helping to explain corporate cultures characterized dysfunctional groupthink or valuable group morale, why delusions flow down the hierarchy, and the emergence of market manias sustained by "new-era" thinking, followed by crashes.

In each of these applications, the institutional and market environment was kept very simple, so as to make clear the commonality of the underlying "Mutually Assured Delusion" principle. Enriching these context-specific features of the model would be quite valuable and permit new applications. This is particularly true in the area of hierarchies and organizations, where richer payoff and information structures should be incorporated, along with greater heterogeneity of interests among agents. Two applications worth examining worth examining in greater detail thus include the spread of organizational corruption (e.g., Anand et al. (2005)), and corporate politics. (e.g. Zald (1998)).

"Fantastic faith" and immunity to evidence are also clearly at work in political ideology. In Bénabou (2008) I thus embed the model into a political-economy setting, to analyze society-wide beliefs concerning the relative merits and proper scope of the state versus the market. A common principle is thus shown to help explain reality distortions in both organizational and political culture. Another application to politics could be to the spread and persistence of conspiracy theories.

A different class of collective delusions which the model so far does not explain are mass panics and hysterias. Understanding the sources and transmissions mechanisms that underlie delusional group pessimism, rather than optimism, remains an interesting open question for further research.

Appendix

In the proofs of Proposition 1 and 2 given below, I maintain the text's focus on cognitive decisions in state L, implicitly fixing everyone's recall strategy in state H to $\lambda_H = 1$. Then, in Lemma 4, I show that this is not a binding restriction: with the payoffs (1), there exists no equilibrium with $\lambda_H < 1$ and no profitable individual deviation to $\lambda_H^i < 1$ from any equilibrium in which $\lambda_H = 1$. These and other results are proved using a more general specification, which is also serves to establish Proposition 7:

$$(A.1) \qquad U_2^i \equiv \theta \left[\alpha e^i + (1 - \alpha) e^{-i} \right] + \gamma,$$

where γ is also state-dependent and $\Delta \gamma \equiv \gamma_H - \gamma_L$ may be of either sign. .

Proof of Proposition 1 i) Let $\equiv \Psi(\lambda^i, s | \lambda^{-i})$ denote the right-hand side of (10). Since it is increasing in λ^i, agent i's optimal awareness strategy is uniquely determined as

a) $\lambda^i = 1$ if $\Psi(1, s | \lambda^{-i}) \leq 0$. By (10), and noting that $\alpha \theta_L + \Delta \theta + (1 - \alpha) \lambda^{-i} \theta_L \geq \min \{\Delta \theta, \theta_H\} > 0$, this means

$$(A.2) \qquad s \leq \frac{m/\delta + c - \delta \alpha \theta_L}{\alpha \theta_L + \Delta \theta + (1 - \alpha) \lambda^{-i} \theta_L} \equiv \underline{s}(\lambda^{-i}).$$

b) $\lambda^i = 0$ if $\Psi(0, s | \lambda^{-i}) \geq 0$. By (10), and noting that $\alpha \theta_L + q \left[\Delta \theta + (1 - \alpha) \lambda^{-i} \theta_L \right] \geq \min \{q \Delta \theta, q \theta_H + (1 - q) \theta_L\} > \min \{q \Delta \theta, c/(s + \delta)\} > 0$, this means

$$(A.3) \qquad s \geq \frac{m/\delta + c - \delta \alpha \theta_L}{\alpha \theta_L + q \left[\Delta \theta + (1 - \alpha) \lambda^{-i} \theta_L \right]} \equiv \bar{s}(\lambda^{-i}).$$

Moreover, $\underline{s}(\lambda^{-i}) < \bar{s}(\lambda^{-i})$, since

$$(A.4) \qquad \Delta \theta + (1 - \alpha) \lambda^{-i} \theta_L \geq \Delta \theta + (1 - \alpha) \lambda^{-i} \min \{\theta_L, 0\}$$
$$\geq \Delta \theta + \min \{\theta_L, 0\} = \min \{\theta_H, \Delta \theta\} > 0.$$

c) $\lambda^i \in (0, 1)$ is the unique solution to $\Psi(\lambda^i, s | \lambda^{-i}) = 0$ for $\Psi(0, s | \lambda^{-i}) < 0 < \Psi(1, s | \lambda^{-i})$, which corresponds to $\underline{s}(\lambda^{-i}) < s < \bar{s}(\lambda^{-i})$.

ii) and iii) follow from the monotonicity properties of Ψ with respect to θ_L and α. Note that assumption of symmetry in strategies was imposed (λ^{-i} could, a priori, be the mean of heterogenous recall rates); therefore, the only equilibria are the symmetric ones described in the proposition. ∎

Proof of Proposition 2 By Proposition 1, $\lambda = 1$ is an equilibrium when $\Psi(1, s | 1) \leq 0$, or

$$(A.5) \qquad s \leq \frac{m/\delta + c - \delta \alpha \theta_L}{\alpha \theta_L + \Delta \theta + (1 - \alpha) \theta_L} = \frac{m/\delta + c - \delta \alpha \theta_L}{\theta_H} \equiv \underline{s}(1),$$

and $\lambda = 0$ is an equilibrium when $\Psi(0, s|0) \geq 0$, or

$$(A.6) \qquad s \geq \frac{m/\delta + c - \delta\alpha\theta_L}{\alpha\theta_L + q\Delta\theta} \equiv \bar{s}(0).$$

Finally, $\lambda \in (0, 1)$ is an equilibrium if and only if $\Psi(\lambda, s|\lambda) = 0$. Now, from (10) and (7),

$$(A.7) \qquad \Psi(\lambda, s|\lambda) = -m/\delta - c + (\delta + s)\alpha\theta_L + sq\left(\frac{\Delta\theta + (1-\alpha)\lambda\theta_L}{q + (1-q)(1-\lambda)}\right).$$

This function is either increasing or decreasing in λ, depending on the sign of $(1-\alpha)\theta_L + (1-q)\Delta\theta$. One can also check, using (A.2)-(A.3), that the same expression governs the sign of $\underline{s}(1) - \bar{s}(0)$. The equilibrium set is therefore determined as follows:

a) If (11) does not hold, $\Psi(\lambda, s|\lambda)$ is increasing, so $\Psi(0, s|0) < \Psi(1, s|1)$, or equivalently $\underline{s}(1) < \bar{s}(0)$ by (A.2)-(A.3). There is then a unique equilibrium, equal to $\lambda = 1$ if $\Psi(1, s|1) \leq 0$, interior if $\Psi(0, s|0) < 0 < \Psi(1, s|1)$, and equal to $\lambda = 0$ if $0 < \Psi(0, s|0)$.

b) If (11) does hold, $\Psi(\lambda, s|\lambda)$ is decreasing, so $\Psi(1, s|1) < \Psi(0, s|0)$, or equivalently $\bar{s}(0) < \underline{s}(1)$ by (A.2)-(A.3), and

$- \lambda = 1$ is the unique equilibrium for $\Psi(0, s|0) \leq 0$, meaning that $s \leq \bar{s}(0)$, while $\lambda = 0$ is the unique equilibrium for $\Psi(1, s|1) \geq 0$, meaning that $s \geq \underline{s}(1)$;

$-$ for $\Psi(1, s|1) < 0 < \Psi(0, s|0)$, or $\bar{s}(0) < s < \underline{s}(1)$, both $\lambda = 1$ and $\lambda = 0$ are equilibria, together with the unique solution to $\Psi(\lambda, s|\lambda) = 0$, which is interior. ∎

Proof of Proposition 3 Following the same steps as in the symmetric case and denoting Λ^{-i} the vector of other agent's strategies, it is easy to show that

$$(A.8) \qquad \underline{s}^i(\Lambda^{-i}) \equiv \frac{m^i/\delta^i + c^i - \delta^i\left(a_L^{ii} - b_L^{ii}\right)}{\Sigma_{j=1}^n\left(a_H^{ji} - a_L^{ji}\right) + \Sigma_{j\neq i}\lambda^j\left(a_L^{ji} - b_L^{ji}\right) + a_L^{ii} - b_L^{ii}},$$

$$(A.9) \qquad \bar{s}^i(\Lambda^{-i}) \equiv \frac{m^i/\delta^i + c^i - \delta^i\left(a_L^{ii} - b_L^{ii}\right)}{q\left[\Sigma_{j=1}^n\left(a_H^{ji} - a_L^{ji}\right) + \Sigma_{j\neq i}\lambda^j\left(a_L^{ji} - b_L^{ji}\right)\right] + a_L^{ii} - b_L^{ii}}.$$

Setting $\lambda^j \equiv 1$ in the first equation and $\lambda^j \equiv 0$ in the second yields the result. I next prove the claims for the case $n = 2$ that follow the proposition and are illustrated in Figure 2. To make things simple, let $m^1 = m^1$, $c^1 = c^2$, $\delta^1 = \delta^2$, $a_H^{11} = a_H^{22}$, $a_L^{11} = a_L^{22}$ and $a_H^{11} - a_L^{11} = a_H^{22} - a_L^{22} \equiv a > 0$; finally, set $b^{ij} = 0$ for all i, j. The asymmetry in roles is then captured by $X \equiv \left(a_H^{12} - a_L^{12}\right)/a > \left(a_H^{21} - a_L^{21}\right)/a \equiv x$ and, especially, $Y \equiv -\left(a_L^{12} - b_L^{12}\right)/a > -\left(a_L^{21} - b_L^{21}\right)/a \equiv y$.

I shall first provide conditions ensuring

$$(A.10) \qquad \bar{s}^2(0) < \underline{s}^1(0) < \underline{s}^1(1) < \bar{s}^1(0) < \bar{s}^1(1) < \underline{s}^2(1),$$

which implies (17). From (A.8)-(A.9), the middle inequality is equivalent to $y < (1-q)(1+x)$, which can always be ensured given $q < 1$. The inequalities $\underline{s}^1(0) < \underline{s}^1(1)$ and $\bar{s}^1(0) < \bar{s}^1(1)$ hold for all $y > 0$ (complementarity). Turning finally to the two outer conditions, we have $\bar{s}^2(0) < \underline{s}^1(0)$ if

$$q\left(a_H^{12} - a_L^{12} + a_H^{22} - a_L^{22}\right) > a_H^{21} - a_L^{21} + a_H^{11} - a_L^{11},$$

or $qX > x + 1 - q$, while $\bar{s}^1(1) < \underline{s}^2(1)$ if

$$q\left[a_H^{21} - a_L^{21} + a_H^{11} - a_L^{11} + a_L^{21} - b_L^{21}\right] > a_H^{12} - a_L^{12} + a_H^{22} - a_L^{22} + a_L^{12} - b_L^{12},$$

or $Y > qy + X - qx + 1 - q$; both are clearly satisfied for X sufficiently larger than x and Y sufficiently larger than X. I can now prove the claims (a)-(c) made in the text.

a) The result follows from the fact that $\bar{s}^2(0) \leq s \leq \underline{s}^2(1)$ and the definitions of these two thresholds in Proposition 1.

b) The same definitions imply that an equilibrium with $(\lambda^1, \lambda^2) = (1, 1)$ (respectively, $(\lambda^1, \lambda^2) = (0, 0)$) exists if and only if $s^2 \leq \underline{s}^2(1)$ and $s^1 \leq \underline{s}^1(1)$ (respectively, $s^2 \geq \bar{s}^2(0)$ and $s^1 \geq \bar{s}^1(0)$), which corresponds to the left (respectively, right) region in Figure 2 In the middle region one must therefore have $\lambda^1 = \lambda_1^*(s^1; \lambda^2), \in (0, 1)$, where λ_1^* is the mixed-strategy best-response characterized in Proposition 1. It is decreasing in s^1 and increasing (respectively increasing) in λ^2 since for $a_L^{21} - b_L^{21} = -ya < 0$.

c) Consider now the boundary loci within the middle region. An equilibrium with $(\lambda^1, \lambda^2) = (\lambda_1^*(s^1; 1), 1)$ exists if and only if $s^1 \in \left[\underline{s}^1(1), \bar{s}^1(1)\right]$ and $s^2 \leq \underline{s}^2(\lambda_1^*(s^1; 1))$. This is a decreasing function of s^1, which declines from $\underline{s}^2(\lambda_1^*(\underline{s}^1(1); 1)) = \underline{s}^2(1)$ at $s^1 = \underline{s}^1(1)$ to $\underline{s}^2(\lambda_1^*(\bar{s}^1(0); 1))$ at $s^1 = \bar{s}^1(0)$; For $\left|a_L^{21} - b_L^{21}\right|/a = y$ small enough, $\lambda_1^*(\bar{s}^1(0); \lambda_2)$ is very insensitive to the value of λ_2, so $\lambda_1^*(\bar{s}^1(0); 1) \approx \lambda_1^*(\bar{s}^1(0); 0)) = 0$, so $\underline{s}^2(\lambda_1^*(\bar{s}^1(0); 1)) \approx \underline{s}^2(0) < \bar{s}^2(0)$. Therefore the curve $\underline{s}^2(\lambda_1^*(s^1; 1))$ cuts the lower boundary of S_2 at a point $s_1 < \bar{s}^1(0)$, as on Figure 2.

Similarly, with $(\lambda^1, \lambda^2) = (\lambda_1^*(s^1; 0), 0)$ exists if and only if $s^1 \in \left[\underline{s}^1(0), \bar{s}^1(0)\right]$ and $s^2 \geq \bar{s}^2(\lambda_1^*(s^1; 0))$. This is a decreasing function of s^1, which declines to $\bar{s}^2(\lambda_1^*(\bar{s}^1(0); 0)) = \bar{s}^2(0)$ at $s^1 = \bar{s}^1(0)$, from $\bar{s}^2(\lambda_1^*(\underline{s}^1(1); 0)$ at $s^1 = \underline{s}^1(1)$; For y small enough, $\lambda_1^*(\underline{s}^1(1); \lambda_2)$ is very insensitive to the value of λ_2, so $\lambda_1^*(\underline{s}^1(1); 0) \approx \lambda_1^*(\underline{s}^1(1); 1) = 1$, so $\bar{s}^2(\lambda_1^*(\underline{s}^1(1); 1)) \approx \bar{s}^2(1) > \underline{s}^2(0)$. Therefore, the curve $\bar{s}^2(\lambda_1^*(s^1; 0))$ cuts the upper boundary of S_2 at a point $s_1 > \underline{s}^1(1)$, as in Figure 2. Finally, for $a_L^{21} - b_L^{21} = 0$,

$$\underline{s}^2(\lambda_1^*(s^1; 1)) = \underline{s}^2(\lambda_1^*(s^1; 0)) < \bar{s}^2(\lambda_1^*(s^1; 0)) = \bar{s}^2(\lambda_1^*(s^1; 1)),$$

since agent 1's behavior is independent of that of agent 2. For y small enough, it remains the case that $\underline{s}^2(\lambda_1^*(s^1; 1)) < \bar{s}^2(\lambda_1^*(s^1; 1))$, by continuity. These properties of the two curves imply that equilibria of the form $(\lambda^1, \lambda^2) = (\lambda_1^*(s^1; 1), 1)$, $(\lambda^1, \lambda^2) = (\lambda_1^*(s^1; 0), 0)$ and $(\lambda^1, \lambda^2) = (\lambda_1^*(s^1; \lambda_2), \lambda_2^*(s^2; \lambda_1))$ exist only in the three respective regions indicated in Figure 2. The equilibrium is therefore unique, except possibly in the middle region where both agents mix.

But since it is unique for $x = y = 0$, by continuity it remains so for x and y small enough. ∎

Proof of Propositions 4 and 5 For $\theta_L < 0$, it is easily seen that

(A.11)
$$s^* \equiv \frac{m/\delta + c - \delta\theta_L}{q\theta_H + (1-q)\theta_L} > \max\left\{\frac{m/\delta + c - \delta\alpha\theta_L}{\alpha\theta_L + q\Delta\theta}, \frac{m/\delta + c - \delta\alpha\theta_L}{\theta_H}\right\} = \max\left\{\bar{s}(0), \underline{s}(1)\right\}.$$

For $\theta_L > 0$, it is easily seen that $s^* < \bar{s}(0)$, but $s^* < \underline{s}(1)$ requires that

$$\frac{m/\delta + c - \delta\alpha\theta_L - \delta(1-\alpha)\theta_L}{\theta_L + q\Delta\theta} < \frac{m/\delta + c - \delta\alpha\theta_L}{\theta_L + \Delta\theta}, \quad \text{or}$$

(A.12)
$$(1-q)\Delta\theta\left[m/\delta + c - \delta\alpha\theta_L\right] < \delta(1-\alpha)\theta_L\theta_H,$$

which can go either way. This finishes to establish Proposition 4. The fist part of Proposition 5 follows from (20). Turning to the second, the difference in average welfare between the $\lambda = 0$ and the $\lambda = 1$ cases (whether as equilibria or through commitment) is

(A.13)
$$q\left(U_{H,D}^* - U_{H,R}^*\right) + (1-q)\left(U_{L,D}^* - U_{L,R}^*\right)$$
$$= -qs(1-q)\Delta\theta + (1-q)\left(-m/\delta - c + (\delta + s)s\theta_L + sq\Delta\theta\right),$$

since $r(1) = 1$ and $r(0) = q$; hence the result. ∎

Proof of Proposition 6 Assume that at $t = 0$, everyone else produces $k^{-i} = K$, and denote the proportions of realists λ^{-i}. Since producing at $t = 1$ (respectively, not producing) is a dominant strategy given posterior $\mu^j = r(\lambda^j) \geq q$ (respectively, $\mu^j = 0$), the price in state L will be $P_L(K + (1 - \lambda^{-i})E)$ and the date-0 expected utilities of realism and denial equal to

$$U_{L,R}(\lambda^i, \lambda^{-i}; k^i)/\delta = (\delta + s)P_L(K + (1 - \lambda^{-i})E)k^i,$$
$$U_{L,D}(\lambda^i, \lambda^{-i}; k^i)/\delta = -m/\delta + (\delta + s)P_L(K + (1 - \lambda^{-i})E)(k^i + E) - cE$$

(A.14)
$$+ sr(\lambda^i)\left[P_H(K + E) - P_L(K + (1 - \lambda^{-i})E)\right](k^i + E)..$$

The net incentive for denial, $\Delta U_L \equiv U_{L,D} - U_{L,R}$, is thus given by

(A.15) $\quad [\Delta U_L(\lambda^i, \lambda^{-i}; \bar{k};) + m]/\delta = \left[(\delta + s)P_L(K + (1 - \lambda^{-i})E) - c\right]E,$
$$+ sr(\lambda^i)\left[P_H(K + E) - P_L(K + (1 - \lambda^{-i})E)\right](k^i + E).$$

Setting $r(\lambda^i) = 1$, realism is a (personal-equilbrium) best response to λ^{-i} for an agent entering period 1 with stock k^i if

(A.16) $\quad m/\delta \geq \left[(\delta + s)P_L(K + (1 - \lambda^{-i})E) - c\right]E$
$$+ s\left[P_H(K + E) - P_L(K + (1 - \lambda^{-i})E)\right](k^i + E).$$

Conversely, denial $(r(\lambda^i) = q)$ is a (personal-equilibrium) best response for i if

(A.17)
$$m/\delta \;\leq\; \left[(\delta + s)P_L(K + (1 - \lambda^{-i})E) - c\right] E$$
$$+ sq \left[P_H(K + E) - P_L(K + (1 - \lambda^{-i})E)\right](k^i + E).$$

For given k^i and λ^{-i}, these two conditions are mutually exclusive. When neither holds, there is a unique $\lambda^i \in (0, 1)$ that equates ΔU_L equal to zero, defining a mixed-strategy (personal equilibrium) best-response. The next step is to solve for (symmetric) social equilibria..

   1. Realism. From (A.16), $\lambda^i = \lambda^{-i} = 1$ is an equilibrium in cognitive strategies if

(A.18)
$$\left[(\delta + s)P_L(K) - c\right] E + s\left[P_H(K + E) - P_L(K)\right](k^i + E) \leq m/\delta.$$

This condition holds for all $k^i \leq K$ if and only if

(A.19)
$$s \leq \frac{m/\delta + [c - \delta P_L(K)]E}{[P_H(K + E) - P_L(K)](K + E) + P_L(K)E} \equiv \underline{s}(1; K).$$

Moving back to the start of period 0, one now verifies that it is indeed an equilibrium for everyone to produce $k^i = K$. Since agents will respond to market signals $\sigma = H, L$, the expected price is $qP_H(K + E) + (1 - q)P_L(K) > 0$, whereas the cost of period-0 production is 0 (more generally, it suffices that it be small enough). Thus, it is optimal to produce to capacity.

   2. Denial equilibrium. From (A.17), $\lambda^i = \lambda^{-i} = 0$ is a cognitive equilibrium if

(A.20)
$$\left[(\delta + s)P_L(K + E) - c\right] E + sq\left[P_H(K + E) - P_L(K + E)\right](k^i + E) \geq m/\delta.$$

This condition holds for $k^i = K$ if

(A.21)
$$s > \frac{m/\delta + [c - \delta P_L(K + E)]E}{q[P_H(K + E) - P_L(K + E)](K + E) + P_L(K + E)E} \equiv \bar{s}(0; q, K).$$

An agent with low k^i, however, has less incentive to engage in denial. In particular, for $s < \underline{s}(1; K)$, (A.18) for $k^i = 0$ precludes (A.20) from holding at $k^i = 0$. Let $\bar{k}(s, q)$ therefore denote the unique solution in k^i to the linear equation

(A.22)
$$\left[(\delta + s)P_L(K + E) - c\right] E + sq\left[P_H(K + E) - P_L(K + E)\right](k^i + E) = m/\delta.$$

Subtracting from (A.22) the equality obtained by evaluating (A.20) at $s = \bar{s}(0; q, K)$ yields

$$sq\left[P_H(K + E) - P_L(K + E)\right](K - \bar{k})$$
$$= (s - \bar{s})P_L(K + E)E + (s - \bar{s})q\left[P_H(K + E) - P_L(K + E)\right](K + E),$$

where the arguments are dropped from $\bar{k}$ and $\bar{s}$ when no confusion results. Thus,

$$(A.24) \quad K - \bar{k} = \left(\frac{s - \bar{s}}{s}\right) \left(\frac{qP_H(K+E) + (1-q)P_L(K+E)}{q\left[P_H(K+E) - P_L(K+E)\right]}E + K\right) > \left(1 - \frac{\bar{s}}{s}\right)(K+E)$$

Note that $\bar{k} \leq K$ (and is thus feasible) if and only if $s \geq \bar{s}$. One can now examine the optimal choice of k^i at $t = 0$, which will be either $k^i = K$ or some $k^i \leq \bar{k}$.

(a) For $k^i > \bar{k}(s,q)$, (A.22) implies that denial is the unique best response to $\lambda^{-i} = 0$, leading agent i to produce $e^i = E$ in both states at $t = 1$. These units and the initial k^i will be sold at the expected price $\bar{P}_q(K+E) \equiv qP_H(K+E) + (1-q)P_L(K+E) > 0$.Therefore, producing up to capacity K in period 0 is optimal among all levels $k^i > \bar{k}(s,q)$, and yields ex-ante utility

$$(A.25) \qquad U_D(0, K, K)/\delta = (\delta + s)\bar{P}_q(K+E)(K+E) - cE - (1-q)m/\delta.$$

(b) For $k^i \leq \bar{k}(q;s)$, on the other hand, agent i's continuation (personal-equilibrium) strategy is some $\lambda^i = \lambda(k^i) \geq 0$: in state L he weakly prefers to be a realist. This leads to

$$(A.26) \quad U(\lambda^i, 0, k^i; K)/\delta = (\delta + s)\bar{P}_q(K+E)\left(k^i + E\right) - cE$$
$$- (1-q)\left\{\left(1 - \lambda^i\right)m/\delta - \lambda^i\left[c - (\delta+s)P_L(K+E)\right]E\right\}.$$

The agent prefers $k^i = K$ (even though it will lead him into denial if state L occurs) to any $k^i \leq \bar{k}(q;s)$ if $U_D(0, K, K) > U(\lambda^i, 0, k^i; K)$, or

$$(\delta + s)\bar{P}_q(K+E)(K - k^i) > (1-q)\lambda^i\left\{m/\delta + \left[c - (\delta+s)P_L(K+E)\right]E\right\}.$$

Using (A.24) and $\lambda^i \leq 1$, it suffices that
(A.27)
$$\left[1 - \bar{s}(0;q,K)/s\right]\bar{P}_q(K+E)(K+E) \geq (1-q)\left\{m/\left[\delta(\delta+s)\right] + \left[c/(\delta+s) - P_L(K+E)\right]E\right\}.$$

Since $\bar{P}_q(K+E)$ tends to $P_H(K+E)$ as q tends to 1, this condition will hold for q close enough to 1, provided $s - \bar{s}(0;q,K)$ remains bounded away from 0. The following two lemmas formalize this and related intuitions. (6).

Lemma 1 *Under (27), there exists $\tilde{q} < 1$ such that, for all $q \in [\tilde{q}, 1]$, $\bar{s}(0;q,K) < \underline{s}(1;K)$.*

Proof: By (A.19)-(A.21), $\bar{s}(0;q,K) < \underline{s}(1;K)$ means that
(A.28)
$$\frac{m/\delta + \left[c - \delta P_L(K+E)\right]E}{q\left[P_H(K+E) - P_L(K+E)\right](K+E) + P_L(K+E)E} < \frac{m/\delta + \left[c - \delta P_L(K)\right]E}{\left[P_H(K+E) - P_L(K)\right](K+E) + P_L(K)E}.$$

If (A.28) holds for $m = 0$, the first denominator must be greater than the second, as $P_L(K+E) <$

31

$P_L(K)$. Therefore, (A.28) holds for all $m \geq 0$ if and only if it holds for $m = 0$, or

$$\frac{c - \delta P_L(K+E)}{c - \delta P_L(K)} < \frac{q\left[P_H(K+E) - P_L(K+E)\right](K+E) + P_L(K+E)E}{P_H(K+E)(K+E) - P_L(K)K}$$
$$= \frac{P_H(K+E)(K+E) - P_L(K+E)K}{P_H(K+E)(K+E) - P_L(K)K} - (1-q)\frac{\left[P_H(K+E) - P_L(K+E)\right](K+E)}{P_H(K+E)(K+E) - P_L(K)K}$$

or:

$$(1-q)\frac{\left[P_H(K+E) - P_L(K+E)\right](K+E)}{P_H(K+E)(K+E) - P_L(K)K} < \frac{P_H(K+E)(K+E) - P_L(K+E)K}{P_H(K+E)(K+E) - P_L(K)K} - \frac{c - \delta P_L(K+E)}{c - \delta P_L(K)}.$$

$$= \frac{P_L(K) - P_L(K+E)}{c - \delta P_L(K)}\left[\frac{cK - \delta P_H(K+E)(K+E)}{P_H(K+E)(K+E) - P_L(K)K}\right].$$

Finally, the condition takes the form

$$1 - q < \left(\frac{cK/(K+E) - \delta P_H(K+E)}{c - \delta P_L(K)}\right)\left(\frac{P_L(K) - P_L(K+E)}{P_H(K+E) - P_L(K+E)}\right).$$

Condition (27) ensures that $cK/(K+E) > \delta P_H(K+E)$, hence the result. $\|$

Lemma 2 *Assume (27). For any $\eta \in (0, 1/2)$ define $s_\eta(0; 1, K) \equiv (1 - \eta)\bar{s}(0; 1, K) + \eta\underline{s}(1; K)$. There exists $q^*(\eta) < 1$ such that, for all $q \in (q^*(\eta), 1]$ condition (A.27) holds for all s in the nonempty interval $S_{2\eta}(K) \equiv (s_{2\eta}(0; 1, K)), \underline{s}(1; K))$.*

Proof: For q close to 1 $\bar{s}(0; q, K)$ is close to $\bar{s}(0; 1, K)$, so there exists $\hat{q}(\eta) \in (\tilde{q}, 1]$ such that, for all $q \in (\hat{q}(\eta), 1]$:

$$\bar{s}(0; q, K) < (1 - \eta)\bar{s}(0; 1, K) + \eta\underline{s}(1; K) \equiv s_\eta(0; 1, K) < \underline{s}(1; K)$$

This implies, for any $s \in S_{2\eta}(K)$:

$$\frac{s - \bar{s}(0; q, K)}{s} > \frac{s_{2\eta}(0; 1, K) - s_\eta(0; 1, K)}{\underline{s}(1; K)} = \eta\left(\frac{\underline{s}(1; K) - \bar{s}(0; 1, K)}{\underline{s}(1; K)}\right) = \eta\left(1 - \frac{\bar{s}(0; 1, K)}{\underline{s}(1; K)}\right)$$

Therefore, condition (A.27) holds provided that

$$(\text{A.29}) \quad 1 - q \leq \eta\left(1 - \frac{\bar{s}(0; 1, K)}{\underline{s}(1; K)}\right)\left(\frac{\bar{P}_q(K+E)(K+E)}{m/\left[\delta(\delta + s)\right] + \left[c/(\delta + s) + \bar{s}(0; 1, K) - P_L(K+E)\right]E}\right),$$

which will be the case for all q in some nonempty subinterval $(q^*(\eta), 1]$ of $(\tilde{q}, 1]$. $\|$

The proof of Proposition 6 concludes by showing that for any $s \in S_{2\eta}(K)$, both $(k^i = K, \lambda^i = 1)$ and $(k^i = K, \lambda^i = 0)$ are equilibria of the two-stage game provided $q \in (q^*(\eta), 1]$.Indeed, for such parameters we have $s \in S_{2\eta}(K) \subset (\bar{s}(0; q, K), \underline{s}(1; K))$, and: (i) for $s < \underline{s}(1; K)$, it was shown that when others play $(k^{-i} = K, \lambda^{-i} = 1)$ agent i finds it optimal to also be a realist

and to produce K in period 0; (ii) for $s > \bar{s}(0; q, K)$, it was shown that when others play $(k^{-i} = K,\ \lambda^{-i} = 0)$ agent i prefers to produce K in period 0, with full knowledge that this will lead him to engage in denial if state L, rather than follow any other (k^i, λ^i) strategy. ■

Lemmas for Proposition 7

Lemma 3 *Let $m > 0$. In any equilibrium such that $\xi^i_H \geq \xi^i_L$ (effort is no less for an agent with date-1 recall $\hat{\sigma}^i = H$ than for one with $\hat{\sigma}^i = L$), it must be that $\lambda^i_H = 1$ or $\lambda^i_L = 1$.*

Proof: Let $U^i_{0,\sigma\sigma'}$ denote agent i's expected utility at $t = 0$ when the state is $\sigma^i = L$ and he encodes it as $\hat{\sigma}^i = \sigma'$, for all $(\sigma, \sigma') \in \{L, H\}^2$. Thus $\sigma = \sigma'$ corresponds to realism and $\sigma' \neq \sigma$ to denial (misinterpreting or misremembering), at a cost of m. Let λ^i_H, λ^i_L denote the probabilities with which $\hat{\sigma}^i = \sigma^i$ in each state. Without generality one can impose $\lambda^i_H + \lambda^i_L \geq 1$ (otherwise, simply switch the labels H and L on the two values which the message $\hat{\sigma}^i$). Let r^i_H and r^i_L, given by

$$(A.30) \qquad r^i_H \equiv \Pr\left[\sigma^i = H \mid \hat{\sigma}^i = H\right] = \frac{q\lambda^i_H}{q\lambda^i_H + (1-q)(1-\lambda_L)},$$

$$(A.31) \qquad r^i_L \equiv \Pr\left[\sigma^i = L \mid \hat{\sigma}^i = L\right] = \frac{(1-q)\lambda_L}{(1-q)\lambda^i_L + q(1-\lambda^i_H)},$$

denote the resulting "reliability" of a message or recollection $\hat{\sigma}^i = H, L$ when the agent's *equilibrium* strategy is λ^i_H, λ^i_L, and note that $r^i_H \geq q \geq 1 - r^i_L$. Finally, let ξ^i_H (resp. ξ^i_L) the probability with which the exerts effort ($e^i = 1$) given $\hat{\sigma}^i = H, L$.

Following the same reasoning as in text, the incentive to misinterpret or misremember a signal H as L (gross of the cost m) is given by

$$(A.32) \quad \left(U^i_{0,HL} - U^i_{0,HH} + m\right)/\delta = s\left(1 - r^i_L - r^i_H\right)(\gamma_H - \gamma_L) + \left(\xi^i_H - \xi^i_L\right)[c - \delta\alpha\theta_H]$$
$$+ s\alpha\left\{\left[(1 - r^i_L)\xi^i_L - r^i_H\xi^i_H\right]\theta_H - \left[(1 - r^i_H)\xi^i_H - r^i_L\xi^i_L\right]\theta_L\right\}$$
$$+ s(1 - \alpha)\left(1 - r^i_L - r^i_H\right)\left\{\left[\lambda^{-i}_H\xi^{-i}_H + (1 - \lambda^{-i}_H)\xi^{-i}_L\right]\theta_H\right.$$
$$\left. - \left[\lambda^{-i}_L\xi^{-i}_L + (1 - \lambda^{-i}_L)\xi^{-i}_H\right]\theta_L\right\}.$$

The incentive to miscode L as H is given by the same expression, with H and L switched:

$$(A.33) \quad \left(U^i_{0,LH} - U^i_{0,LL} + m\right)/\delta = s\left(1 - r^i_H - r^i_L\right)(\gamma_L - \gamma_H) + \left(\xi^i_L - \xi^i_H\right)[c - \delta\alpha\theta_L]$$
$$+ s\alpha\left\{\left[(1 - r^i_H)\xi^i_H - r^i_L\xi^i_L\right]\theta_L - \left[(1 - r^i_L)\xi^i_L - r^i_H\xi^i_H\right]\theta_H\right\}$$
$$+ s(1 - \alpha)\left(1 - r^i_H - r^i_L\right)\left\{\left[\lambda^{-i}_L\xi^{-i}_L + (1 - \lambda^{-i}_L)\xi^{-i}_H\right]\theta_L\right.$$
$$\left. - \left[\lambda^{-i}_H\xi^{-i}_H + (1 - \lambda^{-i}_H)\xi^{-i}_L\right]\theta_H\right\}.$$

This implies:

$$(A.34) \qquad \left(U^i_{0,HL} - U^i_{0,HH}\right)/\delta + \left(U^i_{0,LH} - U^i_{0,LL}\right)/\delta = -2m/\delta - \left(\xi^i_H - \xi^i_L\right)\delta\Delta\theta.$$

Therefore, as long as $2m/\delta + \left(\xi_H^i - \xi_L^i\right)\delta\Delta\theta > 0$, one of the two terms on the left-hand side must be strictly negative, implying that $\lambda_H^i = 1$ or $\lambda_L^i = 1$. $\blacksquare$

Lemma 4 *1) For $\Delta\gamma \geq -(1-\alpha)\min\{\theta_H, \Delta\theta\}$ there can be no equilibrium with $\lambda_H = 0$, and no profitable individual deviation to $\lambda_H^i < 1$ from any equilibrium in which $\lambda_H = 1$.*

2) For $\Delta\gamma > -\min\left\{(1-\alpha)\theta_H, (1-\alpha)\Delta\theta, \kappa^*(s)\Delta\theta)\right\}$, where $\kappa^*(s) > 0$ is given by (A.36) below, there can be no equilibrium with $\lambda_H < 1$. Thus, the results of Propositions 2-5 remain unchanged, up to the substitution of $\Delta\gamma + \Delta\theta$ for $\Delta\theta$ everywhere.

Proof. Note first that, given (4), the fact that $r_H^i \geq q$ implies $\xi_H^i = 1$, so $\lambda_H^i = 1$ or $\lambda_L^i = 1$ by Lemma 3. Moreover, in (symmetric) a equilibrium, $\xi_H^{-i} = \xi_H^i = 1$.

1. *Equilibria with $\lambda_H = 1$.* This implies $r_L^i = 1$, so $\xi_L^i = 0 = \xi_L^{-i}$ and (A.32) becomes

$$
\begin{aligned}
\left(U_{0,HL}^{i,} - U_{0,HH}^{i,} + m\right)/\delta &= -sr_H^i\left(\gamma_H - \gamma_L\right) + [c - \delta\alpha\theta_H] \\
&\quad - s\alpha\left[r_H^i\theta_H + \left(1 - r_H^i\right)\theta_L\right] - sr_H^i\left(1 - \alpha\right)\left[\theta_H - \left(1 - \lambda_L^{-i}\right)\theta_L\right] \\
&= -[(\delta + s)\alpha(r_H^i\theta_H + (1 - r_H^i)\theta_L) - c] - sr_H^i\Delta\gamma \\
&\quad - \Delta\theta[\delta\alpha(1 - r_H^i) + sr_H^i(1 - \alpha)] - sr_H^i\left(1 - \alpha\right)\lambda_L^{-i}\theta_L.
\end{aligned}
$$

The first term is negative since $r_H^i \geq q$, so it suffices that

$$
sr_H^i\Delta\gamma \geq -\Delta\theta[\delta\alpha(1 - r_H^i) + sr_H^i(1 - \alpha)] - sr_H^i\left(1 - \alpha\right)\lambda_L^{-i}\theta_L.
$$

This inequality is linear in r_H^i and clearly holds for $r_H^i = 0$. For $r_H^i = 1$, it takes the form $\Delta\gamma \geq -(1-\alpha)\left[\Delta\theta + \lambda_L^{-i}\theta_L\right]$, which holds whatever the sign of θ_L when $\Delta\gamma \geq -(1-\alpha)\min\{\Delta\theta, \theta_H\}$. Thus, an individual deviation to miscoding H as L is never profitable. As to miscoding L as H, (A.33) becomes

$$
\begin{aligned}
\left(U_{0,LH}^{i,} - U_{0,LL}^{i,} + m\right)/\delta &= -[c - \delta\alpha\theta_L] + s\alpha\left[\left(1 - r_H^i\right)\theta_L + r_H^i\theta_H\right] \\
&\quad + s\left(1 - \alpha\right)r_H^i\left[\theta_H - \left(1 - \lambda_L^{-i}\right)\theta_L\right] + sr_H^i\left(\gamma_H - \gamma_L\right) \\
&= -[c - (\delta + s)\alpha\theta_L] + sr_H^i\left[\Delta\theta + \Delta\gamma + (1 - \alpha)\lambda_L^{-i}\theta_L\right],
\end{aligned}
$$

which is identical to (10) except that $\Delta\theta$ is replaced by $\Delta\theta + \Delta\gamma$. Therefore, all the previous results and formulas shown for $\Delta\gamma = 0$ and imposing $\lambda_H^i \equiv 1$ remain the same, provided $\Delta\theta + \Delta\gamma >$ replaces $\Delta\theta$ wherever it appears.

2. *Ruling out equilibria with $\lambda_H < 1 = \lambda_L$.* If $\lambda_H^i < 1$ then $\lambda_L^i = 1$ by the Lemma 3, so $r_H^i = 1$ and hence $\xi_H^i = 1 = \xi_H^{-i}$. Therefore, (A.32) simplifies to:

$$
\begin{aligned}
\left(U_{0,HL}^{i,} - U_{0,HH}^{i,} + m\right)/\delta &= -\left(1 - \xi_L^i\right)\left[(\delta + s)\alpha\theta_H - c\right] \\
&\quad - sr_L^i\left\{\Delta\theta\left[\alpha\xi_L^i + (1 - \alpha)\xi_L^{-i}\right] + \Delta\gamma + (1 - \alpha)\lambda_H^{-i}\left(1 - \xi_L^{-i}\right)\theta_H\right\}.
\end{aligned}
$$

In (symmetric) equilibrium $\xi_L^i = \xi_L^i$ and $\lambda_H^i = \lambda_H^{-i}$, so this expression is strictly negative and no equilibrium with $\lambda_H^i < 1$ exists, when

$$(A.35) \qquad \xi_L^i \Delta\theta + \left(1 - \xi_L^i\right) \lambda_H^i (1 - \alpha) \theta_H + \Delta\gamma \geq 0$$

For $\Delta\theta + \Delta\gamma \geq 0$, we can rule out any equilibrium with $\xi_L^i = 1$, and in particular any equilibrium with $\lambda_H^i = 0$ (which implies $r_L^i = 1 - q$, so $\xi_L^i = 1$). As to an equilibrium with $\xi_L^i < 1$, given $\lambda_L^i = 1$ this requires that λ_H^i not be below the critical value that makes an agent indifferent to working or not, given $\hat\sigma^i = L : \theta_L + [1 - r_L(\lambda_H, 1)]\Delta\theta \leq c/\alpha(s + \delta)$, or

$$(A.36) \qquad \lambda_H^i (1 - \alpha) \left(\frac{\theta_H}{\Delta\theta}\right) \geq (1 - \alpha) \left(\frac{\theta_H}{\Delta\theta}\right) \left[1 - \left(\frac{1-q}{q}\right) \left(\frac{c/\alpha(s+\delta) - \theta_L}{\theta_H - c/\alpha(s+\delta)}\right)\right] \equiv \kappa^*(s).$$

Therefore, by (A.35), any equilibrium with $\xi_L^i < 1$ is ruled out for $\Delta\gamma \geq -\Delta\theta \min\{1, \kappa^*(s)\}$.

Note finally, that since $\kappa^*(s)$ is increasing, if the second inequality in (4) is strengthened to

$$(A.37) \qquad q\theta_H + (1 - q)\theta_L > c/\alpha\delta,$$

then $\kappa_H^*(0) > 0$ and such equilibria are ruled out for *any* s if $\Delta\theta \min\{1, \kappa^*(0)\} + \Delta\gamma > 0$. ∎

Proof of Proposition 7 I again show the result for the more general specification (A.1), under which $\kappa \geq \max\{1, \theta_H/\Delta\theta\}$ is a special case of $\Delta\gamma \leq -\max\{\Delta\theta, \theta_H\}$.

Note first that since $1 - r_L^i \leq q$, (29) implies that $\xi_L^i = 0$, hence again Lemma 3 shows that $\left(1 - \lambda_H^i\right)\left(1 - \lambda_L^i\right) = 0$. Moreover, in (symmetric) a equilibrium, $\xi_L^{-i} = \xi_L^i = 0$.

1. *Ruling out equilibria with $\lambda_L^i < 1 = \lambda_H^i$.* If $\lambda_L^i < 1$ then $\lambda_H^i = 1 = \lambda_H^{-i}$ in equilibrium by the Lemma 3 and symmetry, so $r_L^i = 1$ and hence $\xi_L^i = 0 = \xi_L^{-i}$. Therefore, (A.33) simplifies to:

$$\left(U_{0,LH}^i - U_{0,LL}^i + m\right)/\delta = sr_H^i\Delta\gamma - \xi_H^i [c - \delta\alpha\theta_L] + s\alpha\xi_H^i \left[\left(1 - r_H^i\right)\theta_L + r_H^i\theta_H\right]$$
$$+ sr_H^i (1 - \alpha)\xi_H^{-i}\left[\lambda_H^{-i}\theta_H - \left(1 - \lambda_L^{-i}\right)\theta_L\right]$$
$$= -\xi_H^i [c - (s + \delta)\alpha\theta_L] + sr_H^i \left\{\Delta\gamma + \xi_H^i\left[\Delta\theta + (1 - \alpha)\lambda_L^i\theta_L\right]\right\}.$$

Since $\Delta\gamma + \xi_H^i \left[\Delta\theta + (1 - \alpha)\lambda_L^{-i}\theta_L\right] \leq \Delta\gamma + \xi_H^i [\Delta\theta + \max\{0, \theta_L\}] < 0$, the previous expression is strictly negative, and no equilibrium with $\lambda_L^i < 1$ exists.

2. *Equilibria with $\lambda_L = 1$.* This implies $r_H^i = 1$ and so $\xi_H^i = 1 = \xi_H^{-i}$ and (A.33) becomes

$$\left(U_{0,LH}^i - U_{0,LL}^i + m\right)/\delta = -sr_L^i (\gamma_L - \gamma_H) - [c - \delta\alpha\theta_L] + s\alpha\theta_H + sr_L^i (1 - \alpha)\lambda_H^{-i}\theta_H$$
$$= -[c - (\delta + s)\alpha(r_L^i\theta_L + (1 - r_L^i)\theta_H)]$$
$$+ sr_L^i\Delta\gamma - (1 - r_L^i)\delta\alpha\Delta\theta + sr_L^i \left[\alpha\Delta\theta + (1 - \alpha)\lambda_H^{-i}\theta_H\right].$$

The first term is negative since $r_L^i \leq 1 - q$,so it suffices that

(A.38) $$sr_L^i \Delta \gamma \leq (1 - r_L^i)\delta\alpha\Delta\theta - sr_L^i \left[\alpha\Delta\theta + (1 - \alpha)\lambda_H^{-i}\theta_H\right].$$

This inequality is linear in r_L^i and clearly holds for $r_L^i = 0$. For $r_L^i = 1$, it takes the form $\Delta\gamma \leq -\left[\alpha\Delta\theta + (1 - \alpha)\lambda_H^{-i}\theta_H\right]$, which holds for all λ_H^i if $\Delta\gamma \leq -\left[\alpha\Delta\theta + (1 - \alpha)\theta_H\right] = -\left[\Delta\theta + (1 - \alpha)\theta_L\right]$. This last expression is greater than $-\max\{\Delta\theta, \theta_H\}$ whatever the sign of θ_L, hence the result. Therefore, no individual deviation to $\lambda_L^i < 1$ can be profitable.

As to (A.32), it is

$$\left(U_{0,HL}^{i,} - U_{0,HH}^{i,} + m\right)/\delta = -sr_L^i\left(\gamma_H - \gamma_L\right) + [c - \delta\alpha\theta_H] - s\alpha\theta_H - s(1 - \alpha)r_L^i\lambda_H^{-i}\theta_H$$
$$= -[(s + \delta)\alpha\theta_H - c] - sr_L^i\left[\Delta\gamma + (1 - \alpha)\lambda_H^{-i}\theta_H\right].$$

Since $-\Delta\gamma - \theta_H > 0$, $\lambda_H^i = 1$ is an equilibrium (implying $r_L^i = 1$) if and only if

(A.39) $$s \leq \frac{m/\delta + \delta\alpha\theta_H - c}{-\Delta\gamma - \theta_H} \equiv \underline{s}(1).$$

Similarly, $\lambda_H^i = 0$ is an equilibrium (implying $r_L^i = 1 - q$) if and only if

(A.40) $$s \geq \frac{m/\delta + \delta\alpha\theta_H - c}{(1 - q)(-\Delta\gamma) - \alpha\theta_H} \equiv \bar{s}(0),$$

if $-\Delta\gamma > \alpha\theta_H/(1 - q)$, otherwise, let $\bar{s}(0) \equiv +\infty$. Multiple equilibria occur for $\bar{s}(0) < \underline{s}(1)$, i.e. $q(-\Delta\gamma) < (1 - \alpha)\theta_H$. The treatment of the mixed-strategy equilibrium is similar to that in Proposition 2. ■

REFERENCES

Acemoglu, Daron, Chernozhukov, Victor and Muhamet Yildiz (2006) "Learning and Disagreement in an Uncertain World," NBER Working Papers12648.

Akerlof, George, and William Dickens (1982) "The Economic Consequences of Cognitive Dissonance," *American Economic Review,* 72, 307—319.

Anand, Vikas, Ashforth, Blake E. and Mahendra Joshi (2005) "Business as Usual: The Acceptance and Perpetuation of Corruption in Organizations," *Academy of Management Executive,* 19(4), 9-23.

Banerjee, Abhijit (1992) "A Simple Model of Herd Behavior," *Quarterly Journal of Economics,* 107(3), 797-817.

Barrienuovo, Alexei (2006) "Warning on Enron is Recalled". *The New York Times,* March 16.

Bénabou, Roland (2008) "Ideology,"*Journal of the European Economic Association,* 6(2-3), 321–352.

Bénabou, Roland, and Jean Tirole (2002) "Self–Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117 , 871–915.

Bénabou, Roland, and Jean Tirole (2004) "Willpower and Personal Rules," *Journal of Political Economy,* 112, 848–887.

Bénabou, Roland, and Jean Tirole (2006a) "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics,* 121(2), May, 699-746.

Bénabou, Roland, and Jean Tirole (2006b) "Incentives and Prosocial Behavior," *American Economic Review,* 96(5), December , 1652-1678.

Bénabou, Roland, and Jean Tirole (2007) "Identity, Dignity and Taboos: Beliefs as Assets," CEPR Discussion Paper no. 6123, February.

Bikhchandani, Sushil , Hirshleifer, David, and Ivo Welch (1998) "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *Journal of Economic Perspectives,* 12(3), 151-170.

Brunnermeier, Markus, and Jonathan Parker (2005) "Optimal Expectations," *American Economic Review,* 90, 1092-1118.

Brunnermeier, Markus, Gollier, Christian and Jonathan Parker (2007) "Optimal Beliefs, Asset Prices and the Preference for Skewed Returns," *American Economics Review Papers and Proceedings,* 97(2), 159-165.

Brunswick, Roger, and Hayes, Gary (2002) "The Enron Cult: 'Groupthink' Caused Everyone Associated With Enron To Contribute to Its Downfall," *The Houston Chronicle,* February.

Bryce, Robert (2002) *Pipe Dreams: Greed, Ego, and the Death of Enron.* New York, NY: PublicAffairs.

Caplin, Andrew, and John Leahy (1994) "Business as Usual, Market Crashes, and Wisdom After the Fact," *American Economic Review,* 84(3), 548-565.

Camerer, Colin, and Malmendier, Ulrike (2007) "Behavioral Economics of Organizations," in *Behavioral Economics and Its Applications,* P. Diamond and H. Vartiainen (eds.), Princeton University Press.

Caplin, Andrew and John Leahy (2001) "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics,* 116, 55–79.

Carrillo, Juan, and Thomas Mariotti (2000), "Strategic Ignorance as a Self-Disciplining Device," *Review of Economic Studies,* 66 , 529–544.

Cialdini, R. (1984) *Influence: The Psychology of Persuasion.* New York, NY: HarperCollins Publishers.

Columbia Accident Investigation Board (2003) *CIAB Final Report,* especially Chapters 6, 7 and 8. Available at http://caib.nasa.gov/.

Dessi, Roberta (2005) "Collective Memory, Identity, and Cultural Investments," IDEI mimeo.

Cohan, J. (2002) " 'I Didn't Know' and 'I Was Only Doing My Job': Has Corporate Governance Careened Out of Control? A Case Study of Enron's Information Myopia". *Journal of Business Ethics,* 40, 275-299.

Di Tella, R., Galiani, S., and E. Schargrodsky, (2007) "The Formation of Beliefs: Evidence from the Allocation of Land Titles to Squatters," *Quarterly Journal of Economics,* 122(1), 209-41.

Eichennwald, Kurt (2003) "Company Man to the End, After All." *The New York Times,* February 9.

Eichennwald, Kurt (2005) *Conspiracy of Fools: A True Story. New York, NY:* BroadwayBooks.

Fang, H., and Moscarini, G. (2005) "Morale Hazard," *Journal of Monetary Economics,* 52(4), 749-778.

Gabaix, Xavier, Krishnamurthy, Arvind and Olivier Vigneron (2007) "Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market", *Journal of Finance,* 62(2), 557-595,

Gervais, S. and Goldtsein, I. (2005) "The Positive Effects of Self-Biased Perceptions in Teams," Duke University mimeo, November.

Goodman, Peter (2007) "This Is the Sound of a Bubble Bursting," *The New York Times,* December 23.

Haslam, Alexander (2004). *Psychology in Organizations: The Social Identity Approach (2nd ed.).* London, UK & Thousand Oaks, CA: Sage.

Hermalin, Benjamin (1998) "An Economic Theory of Leadership: Leading by Example," *The American Economic Review,* 88(5), 1188-1206.

Hersh, Seymour (2004) *Chain of Command.* New York, NY: HarperCollins Publishers.

Huseman, Richard, and Driver, Russel (1979) "Groupthink: Implications for Small Group Decision Making in Business," in *Readings in Organizational Behavior: Dimensions of Management Action,* R. Richard Huseman and Archie Carral, eds.. Boston, MA: Allyn and Bacon.

Isikoff, Michael, and Corn, David (2007) *Hubris.* New York, NY: Three Rivers Press.

Janis, I. (1972) *Victims of Groupthink: Psychological Studies of Policy Decisions and Fiascoes.* Boston, MA: Houghton Mifflin Company.

Jost, J., and Major, B. (Eds.) (2001). *The Psychology of Legitimacy: Emerging Perspectives on Ideology, Justice, and Intergroup Relations.* New York: Cambridge University Press.

Kaplan, Fred (2008) *Daydream Believers: How a Few Grand Ideas Wrecked America's Power.* New York, NY: John Wiley and Sons.

Karlsson, N., Loewenstein, G. and D. Seppi (2005) "The 'Ostrich Effect': Selective Attention to Information about Investments," Carnegie Mellon University mimeo, May.

Köszegi, Botond (2006) "Emotional Agency," *Quarterly Journal of Economics,* 21(1), 121-156.

Köszegi, Botond (2007) "Utility from Anticipation and Personal Equilibrium," UC Berkeley mimeo, August.

Kuran, T. (1993) "The Unthinkable and the Unthought," *Rationality and Society,* 5, 473-505.

Landier, Augustin, "Wishful Thinking: A Model of Optimal Reality Denial," Massachusetts Institute of Technology mimeo, 2000.

Landier, Augustin, Srarer, David and David Thesmar (2006) "Optimal Dissent in Organizations," NYU mimeo, May.

Lerner, Melvin J. (1982), *The Belief in a Just World: A Fundamental Delusion* (New York, NY: Plenum Press.

Malmendier, Ulrike, and Geoffrey Tate (2008) "Who Makes Acquisitions? CEO Overconfidence and the Market's Reaction," *Journal of Financial Economics,* forthcoming.

Malmendier, Ulrike, and Geoffrey Tate (2005) "CEO Overconfidence and Corporate Investment," *Journal of Finance,* 60 (6), 2661-2700.

McCarthy, Jonathan, and Richard Peach (2004) "Are Home Prices the Next "Bubble"? *New York FED Economic Policy Review,* December, 10(3).

Norris, Floyd (2008) "Color-Blind Merrill in a Sea of Red Flags." The New York Times, May 16.

North, Douglass C. (1990) *Institutions, Institutional Change and Economic Performance.* Cambridge, UK: Cambridge University Press.

Pearlstein, Steven (2006) "Years of Self-Deception Killed Enron and Lay," *The Washington Post,* July 8th.

Prendergast, Canice (1993) "A Theory of 'Yes Men'," *American Economic Review,* 83(4), 757-770.

Reilly, David (2007) "Marking Down Wall Street." *The Wall Street Journal,* September 14, C1.

Reis, R. (2006) "Inattentive Consumers," *Journal of Monetary Economics,* 53 (8), 1761-1800.

Rogers Commission (1986). *Report of the Presidential Commission on the Space Shuttle Challenger Accident.* http://history.nasa.gov/rogersrep/genindex.htm.

Rotemberg, Julio, and Saloner, Garth (2000) "Visionaries, Managers, and Strategic Direction," *Rand Journal of Economics* 31, Winter, 693-716.

Samuelson, Robert (2001) "Enron's Creative Obscurity." *The Washington Post,* December 19.

Schelling, Thomas (1986) The Mind as a Consuming Organ," in David E. Bell, Howard Raiffa, and Amos Tversky, eds., *Decision Making : Discriptive, Normative, and Prescriptive Interactions.* Cambridge, MA: Cambridge University Press.

Scheinkman, Jose, and Wei Xiong (2003) "Overconfidence and Speculative Bubbles," *Journal of Political Economy* 111, 1183-1219.

Shiller, Robert (2003) "From Efficient Markets Theory to Behavioral Finance," *Journal of Economic Perspectives,* 17(1), 83-104

Shiller, Robert (2005) *Irrational Exuberance.* Second Edition, Princeton, NJ: Princeton University Press..

Sims, C. "Implications of Rational Inattention", *Journal of Monetary Economics,* 50(3), 665–690.

Sims, R. (1992) "Linking Groupthink to Unethical Behaviors in Organizations," *Journal of Business Ethics,* 11, 651-662.

Slovic, P. (2007) "If I Look at the Mass I will Never Act: Psychic Numbing and Genocide," *Judgment and Decision-Making,* 2(2), 79-95.

Small, D., Loewenstein, G. and Slovic, P. (2007) "Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims," *Organizational Behavior and Human Decision Processes,* 143–153

Suskind, Ron (2004) "Without a Doubt," *The New York Times,* October 17.

Tabellini, G. (2007) "The Scope of Cooperation: Norms and Incentives," Università Bocconi mimeo, February.

Tenbrusel, Ann, and David Messick (2004) "Ethical Fading: The Role of Self-Deception in Unethical Behavior," *Social Justice Research,* 17(2), 223-662.

Van den Steen, E. (2005) "On the Origins of Shared Beliefs (and Corporate Culture)," MIT Sloan School Working Paper, August.

Zald, Mayer, and Berger, Michael (1978) "Social; Movements in Organizations: Coup d'Etat, Insurgency, and Mass Movements," *The American Journal of Sociology,* 83(4), 823-861.